

University of Groningen

## Expressions of epistemic modality in Mainland Scandinavian

Beijering, Karin

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2012

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Beijering, K. (2012). *Expressions of epistemic modality in Mainland Scandinavian*. s.n.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# **Expressions of epistemic modality in Mainland Scandinavian**

A study into the lexicalization-grammaticalization-pragmaticalization interface

KARIN BEIJERING



**university of  
 groningen**

faculty of arts



The research reported on in this thesis has been carried out under the auspices of the Netherlands National Graduate School of Linguistics (LOT – Landelijke Onderzoekschool Taalwetenschap) and the Center for Language and Cognition Groningen (CLCG) of the Faculty of Arts of the University of Groningen. Publication of this dissertation was financially supported by the University of Groningen.



Groningen Dissertations in Linguistics 106  
ISSN: 0928-0030

ISBN: 978-90-367-5727-0  
ISBN: 978-90-367-5726-3 (electronic version)

© 2012, Karin Beijering

Cover image by Kirsten de Valk, © 2012 - [www.kirstendevalk.nl](http://www.kirstendevalk.nl)

Printed by Wöhrmann Print Service, Zutphen, The Netherlands

RIJKSUNIVERSITEIT GRONINGEN

**Expressions of epistemic modality in Mainland Scandinavian**

A study into the lexicalization-grammaticalization-pragmaticalization interface

**Proefschrift**

ter verkrijging van het doctoraat in de  
Letteren  
aan de Rijksuniversiteit Groningen  
op gezag van de  
Rector Magnificus, dr. E. Sterken,  
in het openbaar te verdedigen op  
donderdag 11 oktober 2012  
om 14:30 uur

door

**Karin Beijering**

geboren op 23 mei 1984  
te Emmen

Promotores	:	Prof. dr. M. Norde Prof. dr. J. Hoeksema
Beoordelingscommissie	:	Prof. dr. O. C. M. Fischer Prof. dr. G. Laureys Prof. dr. J. Nuyts

## Dankwoord

Het schrijven van een proefschrift bestaat uit pieken en dalen die elkaar voortdurend afwisselen. Terugkijkend op de afgelopen vier jaar denk ik met veel plezier terug aan de vele (inter)nationale congressen, workshops en zomer- en winterscholen die ik heb bezocht. Maar er waren ook periodes van flinke tegenslag en overmacht die het me soms erg moeilijk hebben gemaakt om de motivatie en aandacht voor het project vast te houden. Nu het proefschrift na zo'n vier en half jaar voltooid is, ben ik blij dat ik het promotietraject succesvol heb weten af te ronden. Dit was uiteraard niet mogelijk geweest zonder de steun en aanmoediging van vele mensen die elk op hun eigen manier bij het project betrokken zijn geweest of belangstelling hebben getoond. Nu is het moment daar om iedereen te bedanken die me geholpen heeft om het onderzoek uit te voeren dat in dit proefschrift beschreven is.

Allereerst wil ik mijn promotor Muriel Norde bedanken voor haar constructieve feedback, het lezen van de vele vorige versies van hoofdstukken, artikelen en presentaties die aan dit proefschrift voorafgaan, de mogelijkheid om mijn eigen artikelen te schrijven en te publiceren, en het vertrouwen om gezamenlijk werk te presenteren op internationale congressen. Ik heb altijd veel vrijheid in de uitvoering en richting van het project gehad, wat essentieel voor de motivatie en voortgang van het project is geweest, maar zonder onze (wekelijkse) besprekingen en het opstellen van een goede planning met concrete deadlines was het proefschrift nu waarschijnlijk nog steeds 'bijna af' geweest. Gaandeweg heeft het project een meer theoretische wending genomen, wat heeft geresulteerd in een ander proefschrift dan het oorspronkelijke plan, maar ik denk dat we zeker tevreden kunnen zijn met de totstandkoming van de clustermethode die hier uit voortkomt.

Graag bedank ik ook mijn tweede promotor Jack Hoeksema voor hulp bij algemeen taalkundige vragen, en het becommentariëren en lezen van het manuscript in relatief korte tijd. Ook bedank ik de leden van de leescommissie, Olga Fischer, Godelieve Laureys en Jan Nuyts, elk met hun expertise op het gebied van grammaticalisatie, modaliteit en/of Scandinavisch taalkunde, voor hun bereidheid de leescommissie te vormen, het lezen van het manuscript en hun voorgenomen aanwezigheid bij de promotie. In het bijzonder dank ik Jan Nuyts voor zijn uitvoerige commentaren, vragen en suggesties die zeker hebben bijgedragen aan een verbeterde eindversie van het manuscript.

Een promotietraject bestaat gelukkig niet alleen uit veel lezen en schrijven op kantoor, maar ook uit het bezoeken en organiseren van congressen en workshops. De afgelopen vier jaar heb ik de mogelijkheid gehad veel te reizen en mijn onderzoek op (inter)nationale bijeenkomsten te presenteren. Ook heb ik gedeeltes van het proefschrift kunnen publiceren als artikel. Anonieme reviewers en vragenstellers uit het publiek zijn

helaas niet te traceren, maar een aantal van deze onbekenden heeft zeer bruikbare commentaren, vragen en suggesties aangedragen.

Het organiseren van de tweedaagse workshop *Current Trends in Grammaticalization Research*, oktober 2009 in Groningen, en het co-redigeren van het gelijknamige special issue van *Language Sciences*, was niet alleen een leerzame, maar vooral een leuke ervaring. Muriel, Alexandra, Caroline, Lily en Gijs bedankt voor de gezelligheid en goede samenwerking! Meer recentelijk was ik co-organisator van de *Workshop on Pragmaticalization* op *New Reflections on Grammaticalisation*, juli 2012 in Edinburgh. Ik dank Muriel voor de co-organisatie en hoop dat de algehele aversie tegen pragmaticalisatie na deze workshop wat is afgenomen.

Verder was ik de afgelopen jaren verantwoordelijk voor de organisatie van LVLC-bijeenkomsten en de Grammaticalization Reading Group. Ik wil iedereen bedanken die voor korte of langere tijd heeft deelgenomen aan de leesgroep. Therese L-T, Alexandra, Muriel, Remco vP, Oscar, Sander, Trevor, Harrie en Mike bedankt voor het bespreken van de grammaticalisatie literatuur en jullie ideeën over verschillende vraagstukken binnen grammaticalisatie studies.

Gedurende een promotietraject is het een komen en gaan van collega's. Vanaf de start van het project heb ik verschillende kantoren, en kantoorgenoten, gehad op de tweede verdieping van het Harmoniegebouw. En nu, na zo'n tien (!) jaar van studie en onderzoek, is het voor mij tijd om de afdeling Scandinavische Talen en Culturen -en de tweede etage- te verlaten.

Koffie- en theepauzes, lunch, gesprekken op de gang of kantoor, en (Scandinavisten)borrels met Oscar, Lena, Roald, Charlotte, Kirsten, Hinka, Petra, Janke, Anja, Ester, Nanna, Tette, Femke, Jelena, Stefanie, Karin H, Therese L, Therese L-T, Sebastian, Nelleke, Lieselotte, Remco vP, Remco K, waren een welkome afwisseling op de soms lange werkdagen. *En stor takk til Lena, Kirsten, Charlotte og Nanna for å lese gjennom mine danske, norske og svenske eksempelsetninger.*

Wyke, Gorus en het secretariaat Cluster Europese talen en Regiostudies bedankt voor jullie hulp met allerlei bureaucratische, administratieve en/of praktische zaken, en het maken van een praatje in de wandelgangen.

Ook mijn collega PhD studenten bij het CLCG, Nynke, Gisi, Diana, Ildikó, Hanneke, Ankelien, Martijn, Dörte, Daniël, Barbara, Jelena P, Tim vdC, Gideon, Çağrı, Sanne, Jacolien, Myrte, Veerle, Ryan, Trevor, Ruggero, Rika, Bregtje, Tim K, Radek, Aysa en Zhenya, hebben bijdragen aan een gezellige en leuke tijd, vooral door middel van niet werkgerelateerde activiteiten, zoals koffiepauzes, praatjes op de gang, kantoor of bij het koffie- of kopieerapparaat, lunch, sport (voetbal en klimmen), (verjaardags)feestjes, poolen, borrels en avonden/nachten uit in de Groninger binnenstad.

Ik ben blij dat twee van mijn beste vriendinnen, Margreet en Kirsten, mijn paranimfen zijn op de grote dag. Bedankt voor jullie hulp met de voorbereidingen en het organiseren van de promotie. Margreet, bedankt voor je hulp bij het organiseren van de festiviteiten - en het nodige vooronderzoek in de Groninger horeca dat we eigenlijk al tijdens onze studententijd hebben verricht. Kirsten, bedankt voor het ontwerpen van de afbeelding op de omslag van 'het papieren monster' dat bij deze bestreden is.

Tot slot rest mij nog mijn ouders, familie, vrienden en kennissen te bedanken voor hun belangstelling in de voortgang van het project maar vooral voor de soms nodige afleiding in de vorm van een praatje, een mailtje, sms, visite, (korte) vakantie, of een kunstzinnig, culinair of sportief uitje. Johanna, Henk, Nico, Ria en Johan, Henny, Egbert, Judith en Matthijs, Albert en Marchje, Tom en Lisanne, Jan-Albert, Reina, familie Zwart, dank jullie wel voor jullie steun, aanmoediging en interesse!

Klijndijk, 20 augustus 2012





## Table of contents

<i>Glossary</i> .....	6
<i>List of Tables</i> .....	7
<i>List of Figures</i> .....	9
<i>Chapters based on published and submitted peer-reviewed papers</i> .....	10
<b>Chapter 1</b> .....	11
1. Introduction .....	11
1.1 Modality .....	12
1.1.1 Dynamic modality .....	13
1.1.2 Deontic modality .....	14
1.1.3 Epistemic modality .....	14
1.1.4 Evidentiality .....	19
1.1.5 Epistemic modality and related domains .....	20
1.2 The Mainland Scandinavian languages .....	21
1.3 The framework of grammaticalization studies .....	23
1.3.1 Approaches to grammaticalization .....	24
1.3.2 Formal approaches to grammaticalization .....	24
1.3.3 Functional approaches to grammaticalization .....	25
1.3.4 The approach to grammaticalization in the present study .....	25
1.4 The present study .....	26
1.4.1 Sources and method .....	26
1.4.2 Aims and scope .....	27
1.4.3 Organization of the study .....	28
<b>Chapter 2</b> .....	30
2. Preliminaries .....	30
2.1 Definitions .....	30
2.1.1 Grammaticalization .....	30
2.1.1.1 Grammaticalization as a (uni)directional change .....	35
2.1.1.2 Mechanisms of grammaticalization .....	37
2.1.1.3 Some examples of grammaticalization .....	40
2.1.1.4 Properties of grammaticalization .....	42
2.1.1.5 Towards a definition of grammaticalization .....	46
2.1.2 Lexicalization .....	49
2.1.2.1 Characterizations of lexicalization .....	50
2.1.2.2 Properties of lexicalization .....	52
2.1.2.3 Some examples of lexicalization .....	53
2.1.2.4 Towards a definition of lexicalization .....	53

2.1.3 Pragmaticalization.....	56
2.1.3.1 Views on pragmaticalization.....	57
2.1.3.2 Properties of pragmaticalization .....	57
2.1.3.3 Some examples of pragmaticalization .....	59
2.1.3.4 Towards a definition of pragmaticalization .....	60
2.1.4 Subjectification and intersubjectification .....	61
2.1.4.1 Langacker.....	62
2.1.4.2 Traugott.....	63
2.1.4.3 Nuyts .....	64
2.1.4.4 (Inter)subjectivity and (inter)subjectification in the present study .....	65
2.2 Interfaces .....	70
2.2.1 The lexicalization-grammaticalization interface.....	71
2.2.2 The grammaticalization-pragmaticalization interface .....	73
2.2.3 The pragmaticalization-lexicalization interface.....	76
2.2.4 The lexicalization-grammaticalization-pragmaticalization interface.....	77
<b>Chapter 3 .....</b>	<b>81</b>
3. Case study I: modal auxiliaries MUST/MAY.....	81
3.1 Introduction.....	81
3.2 Modals.....	83
3.3 Mainland Scandinavian MUST/MAY .....	84
3.3.1 Etymology of Mainland Scandinavian MUST/MAY .....	87
3.3.2 Development of Mainland Scandinavian MUST/MAY .....	87
3.4 Sources and method .....	91
3.5 Results.....	93
3.5.1 Results for <i>må</i> (present tense).....	93
3.5.1.1 Deontic meanings.....	94
3.5.1.2 Epistemic meanings .....	95
3.5.1.3 Optative meanings .....	96
3.5.1.4 Concessive meanings .....	97
3.5.1.5 Overall results .....	98
3.5.2 Results for <i>måtte</i> (past tense).....	98
3.5.2.1 Deontic meanings.....	99
3.5.2.2 Epistemic meanings .....	100
3.5.2.3 Optative meanings .....	101
3.5.2.4 Concessive meanings .....	101
3.5.2.5 Eventuality meanings.....	102
3.5.2.6 Overall results .....	103
3.6 Grammaticalization and (inter)subjectification of MUST/MAY .....	103
3.6.1 Grammaticalization of MUST/MAY .....	103

3.6.1.1 Mechanisms in the grammaticalization of MUST/MAY .....	104
3.6.1.2 Primitive changes in the grammaticalization of MUST/MAY .....	105
3.6.1.3 Side effects in the grammaticalization of MUST/MAY .....	107
3.6.1.4 Grammatical status and degrees of grammaticalization for MUST/MAY .....	109
3.6.1.5 (Inter)subjectification of MUST/MAY .....	110
3.7 Summary, discussion and conclusions .....	110
<b>Chapter 4</b> .....	113
4. Case study II: Modal sentence adverbs MON and MAYBE .....	113
4.1 Introduction .....	113
4.2 Adverbs .....	114
4.2.1 Epistemic adverbs .....	115
4.2.2 Epistemic adverbs in the Mainland Scandinavian languages .....	117
4.2.2.1 Formal properties of Mainland Scandinavian MAYBE .....	120
4.2.2.2 Formal properties of Mainland Scandinavian MON .....	121
4.3 Case study IIa: Mainland Scandinavian MON .....	123
4.3.1 Semantic properties of MON and MONNE .....	125
4.3.1.1 Danish <i>monne</i> and <i>mon</i> .....	125
4.3.1.2 Norwegian <i>monne</i> and <i>mon</i> .....	126
4.3.1.3 Swedish <i>månde</i> and <i>månne</i> .....	127
4.3.2 Method and sources .....	127
4.3.3 Results .....	128
4.3.3.1 Danish <i>mon</i> .....	129
4.3.3.2 Norwegian <i>mon</i> .....	130
4.3.3.3 Swedish <i>månne</i> .....	130
4.3.3.4 Overall results .....	131
4.3.4 Mainland Scandinavian MON: a case of grammaticalization? .....	134
4.3.4.1 Mechanisms in the development of MON .....	134
4.3.4.2 Primitive changes in the development of MON .....	135
4.3.4.3 Side effects in the development of MON .....	136
4.3.4.4 Linguistic status and degrees of X-ization for MON .....	138
4.4 Case study IIb: Mainland Scandinavian MAYBE .....	139
4.4.1 Semantic properties of MAYBE .....	141
4.4.1.1 Danish <i>måske</i> .....	141
4.4.1.2 Norwegian <i>kanskje</i> .....	142
4.4.1.3 Swedish <i>kanske</i> .....	142
4.4.2 Sources and method .....	143
4.4.3 Results .....	146
4.4.3.1 Clause type I .....	148
4.4.3.2 Clause type II .....	150

4.4.3.3 Clause type III.....	153
4.4.4 Mainland Scandinavian MAYBE: a case of lexicalization, grammaticalization or pragmaticalization?.....	154
4.4.4.1 Mechanisms in the development of MAYBE.....	154
4.4.4.2 Primitive changes in the development of MAYBE.....	155
4.4.4.3 Side effects in the development of MAYBE.....	157
4.4.4.4 Linguistic status and degrees of X-ization for MAYBE.....	159
4.5 Overall summary, discussion and conclusions.....	160
<b>Chapter 5.....</b>	<b>164</b>
5. Case study III: Mental state predicates TROR.....	164
5.1 Introduction.....	164
5.2 Discourse markers.....	165
5.2.1 Mental state predicates of the <i>I think</i> -type.....	166
5.2.1.1 Formal properties of <i>I think</i> .....	166
5.2.1.2 Semantic properties of <i>I think</i> .....	167
5.2.1.3 Development of <i>I think</i> .....	168
5.2.2 Mainland Scandinavian TRO and TROR.....	171
5.2.2.1 Danish <i>tro</i> .....	172
5.2.2.2 Norwegian <i>tro</i> .....	173
5.2.2.3 Swedish <i>tro</i> .....	174
5.2.2.4 Mainland Scandinavian TROR.....	175
5.3 Sources and method.....	178
5.4 Results.....	179
5.4.1 Qualificational and non-qualificational meanings.....	179
5.4.2 Person.....	181
5.5.3 <i>That</i> -deletion.....	183
5.5.5 Position.....	185
5.4.7 Overall results.....	188
5.5 Mainland Scandinavian TROR: lexicalization, grammaticalization or pragmaticalization?.....	189
5.5.1 Mechanisms in the development of TROR.....	189
5.5.2 Primitive changes in the development of TROR.....	190
5.5.3 Side effects in the development of TROR.....	191
5.5.4 Linguistic status and degrees of X-ization for TROR.....	192
5.6 Summary, discussion and conclusions.....	193

<b>Chapter 6.</b>	195
6. Discussion and conclusions	195
6.1 A clustering approach to epistemic expressions at the lexicalization-grammaticalization-pragmaticalization interface	196
6.1.1 Mechanisms in the development of MUST/MAY, MON, MAYBE and TROR	197
6.1.2 Primitive changes in the development of MUST/MAY, MON, MAYBE and TROR	198
6.1.2.1 Syntax	200
6.1.2.1 Morphology	201
6.1.2.3 Phonology/phonetics	201
6.1.2.4 Semantics	202
6.1.2.5 Discourse/pragmatics	202
6.1.3 Side effects in the development of MUST/MAY, MON, MAYBE and TROR	204
6.2 Conclusions	205
6.3 Outlook and suggestions for further research	206
<b>Appendices</b>	208
<b>References</b>	212
<b>Summary</b>	224
<b>Samenvatting</b>	226
<b>Groningen dissertations in linguistics (grodil)</b>	228

## Glossary

### *Corpora*

KDK	KorpusDK
NAK	Norsk Aviskorpus
SK	Språkbankens konkordanser
P95-98	Press 95-98
BII	Bonniersromaner II

### *Dictionaries*

SAOB	Svenska Akademiens Ordbok [Dictionary of the Swedish Academy]
SAOL	Svenska Akademiens Ordlista [Word list of the Swedish Academy]
BO	Bokmålsordboka [The Bokmål dictionary]
DDO	Den Danske Ordbog [The Danish dictionary]
ODS	Ordbog over det Danske Sprog [Dictionary of the Danish language]
VD	Van Dale Groot woordenboek van de Nederlandse taal [Van Dale big dictionary of the Dutch language]
OED	Oxford English Dictionary
EWS	Etymologisches Wörterbuch der deutschen Sprache (Kluge) [Etymological dictionary of the German language]
NSSO	Norstedts Stora Svenska Ordbok [Nordstedt big Swedish dictionary]
PNO	Politikens Nudansk ordbog [Politiken's Danish dictionary]

### *Linguistic terms*

ADJ	adjective	-3SG	third person singular
ADV	adverb	-DAT	dative
AUX	auxiliary	-PAS	passive
MHG	Middle High German	-PAST	past tense
N	noun	-PRF	perfect
OE	Old English		
PDE	Present-day English		
V	verb		
VP	verb phrase		

## List of Tables

Table 1. Epistemic scale for adjectives, nouns, adverbs, auxiliaries and mental state predicates.	18
Table 2. Language situation in Scandinavia during the Viking age ( $\pm$ 800).	21
Table 3. Language situation in Scandinavia around the year 1200.	21
Table 4. Language situation in Scandinavia after the year 1500.	22
Table 5. Language situation in present-day Scandinavia.	22
Table 6. Mechanisms, primitive changes and side effects of grammaticalization.	48
Table 7. Mechanisms, primitive changes and side effects of lexicalization.	55
Table 8. Mechanisms, primitive changes and side effects of pragmaticalization.	61
Table 9. Types of subjectification.	70
Table 10. Converging and diverging properties of lexicalization, grammaticalization and pragmaticalization.	79
Table 11. Inflectional paradigms for <i>må</i> , <i>måtte</i> and <i>måste</i> .	86
Table 12. Mechanisms in the grammaticalization of MUST/MAY.	104
Table 13. Primitive changes in the grammaticalization of MUST/MAY.	105
Table 14. Side effects of grammaticalization in MUST/MAY.	107
Table 15. Types of subjectification in MUST/MAY.	110
Table 16. Frequencies for derived and unverbated epistemic adverbs in the Mainland Scandinavian languages.	118
Table 17. V2 in the Mainland Scandinavian languages.	119
Table 18. Differences in word order between main and subordinate clauses in the Mainland Scandinavian languages.	120
Table 19. Syntactic positions for sentence adverbs in Mainland Scandinavian.	120
Table 20. The peculiar syntactic status of Swedish <i>månne</i> (1).	122
Table 21. The peculiar syntactic status of Swedish <i>månne</i> (2).	122
Table 22. The reanalysis from auxiliary form to interrogative adverb for MON.	124
Table 23. The syntactic distribution of Danish <i>mon</i> .	132
Table 24. The syntactic distribution of Norwegian <i>mon</i> .	133
Table 25. The syntactic distribution of Swedish <i>månne</i> .	133
Table 26. Mechanisms in the development of MON.	135
Table 27. Primitive changes in the development of MON.	136
Table 28. Side effects in the development of MON.	138
Table 29. Types of subjectification in the development of MON.	138
Table 30. Mechanisms in the development of MAYBE.	155
Table 31. Primitive changes in the development of MAYBE.	156
Table 32. Side effects in the development of MAYBE.	157
Table 33. Types of subjectification in the development of MAYBE.	160
Table 34. Mechanisms in the development of MON and MAYBE.	162
Table 35. Primitive changes in the development of MON and MAYBE.	162
Table 36. Side effects in the development of MON and MAYBE.	163



Table 37. Inflectional paradigms for Mainland Scandinavian TRO. ....	172
Table 38. The structural representation of MC and CTP status for TROR.....	176
Table 39. The structural representation of CTP and DM status for TROR. ....	176
Table 40. Positions for sentence adverbs: MAYBE vs TROR .....	177
Table 41. Positions for speech-act adverbials: HONESTLY vs TROR .....	178
Table 42. Mechanisms in the development of TROR.....	190
Table 43. Primitive changes in the development of TROR .....	191
Table 44. Side effects in the development of TROR. ....	192
Table 45. Types of subjectification in the development of TROR .....	193
Table 46. Mechanisms of change in standard cases of X-ization and the case studies. ....	198
Table 47. Primitive changes in standard cases of X-ization and the case studies. ....	200
Table 48. Types of subjectification in standard cases of X-ization and the case studies. ....	203
Table 49. Side effects in standard cases of X-ization and the case studies. ....	204

## List of Figures

Figure 1. Grammaticalization vs Pragmaticalization (Aijmer).....	58
Figure 2. The hierarchy of qualificational categories (Nuyts) .....	65
Figure 3. Synchronic and diachronic interfaces.....	70
Figure 4. Cross-linguistic correlations for premodal, modal and postmodal meanings for MUST/MAY .....	88
Figure 5. Bar chart of the semantic distribution of <i>må</i> (present tense) and <i>måste</i> .....	94
Figure 6. Bar chart of the semantic distribution of <i>mätte</i> (past tense) and <i>måste</i> .....	99
Figure 7. Bar chart of the different meanings and uses of MON .....	129
Figure 8. Bar chart of the distribution of clause type I, II and III for MAYBE.....	148
Figure 9. Bar chart of the distribution for (non-)qualificational meanings of TROR.....	180
Figure 10. Bar chart of the distribution for first, second and third person subjects with TROR.	181
Figure 11. Bar chart of the distribution for <i>that</i> -complementation and <i>that</i> -deletion for TROR	184
Figure 12. Bar chart of the distribution for clause-initial, -medial and -final TROR .....	188

## Chapters based on published and submitted peer-reviewed papers

- **Chapter 3 – Case study I: Modal auxiliaries MUST/MAY**

Beijering, Karin. 2011. Semantic change and grammaticalization: The development of modal and postmodal meanings in Mainland Scandinavian *må*, *måtte* and *måste*. *Nordic Journal of Linguistics* 34(2), 105–132.

Beijering, Karin. (submitted). Grammaticalization and (inter)subjectification: The development of the Swedish modals *må* and *måtte*.

- **Chapter 4 – Case study II: Modal sentence adverbs MON and MAYBE**

Beijering, Karin. 2010. The grammaticalization of Mainland Scandinavian MAYBE. In Bugge, Edit & Lidun Hareide (eds) *Seven mountains Seven Voices* (=Bergen Language and Linguistics Studies, volume 1). Bergen: University of Bergen.  
<https://bells.uib.no/index.php/bells/issue/view/20>

Beijering, Karin. (forthcoming). From modal auxiliary to adverb: the development of Mainland Scandinavian MONNE/MON. In *Festschrift NN*.

# Chapter 1

## 1. Introduction

This thesis investigates the rise of epistemic expressions, i.e. likelihood evaluations on behalf of the speaker, in relation to different types of language change, viz. lexicalization, grammaticalization and pragmaticalization. Developments in the domain of epistemic modality touch upon contentious issues and much-debated topics within grammaticalization studies, such as the lexicalization-grammaticalization interface, the relation between grammaticalization and (inter)subjectification and the status of pragmaticalization. Epistemic modality may, at a first glance, seem an abstract theoretical notion, but it is more basic to everyday life than it at first sight appears to be, as described by Nuyts (2001:xvi) in (1).

- (1) Evaluating the likelihood of a state of affairs is not just a linguistic category, however. It relates directly to the way we perceive, memorize, and act in the physical and social world we live in. More specifically, it taps a crucial dimension of our mental activities, viz. our capacity to reflect on our knowledge and our reasoning with it, i.e. our capacity to metarepresent.

In the present study, four case studies from the domain of epistemic modality in the Mainland Scandinavian languages, i.e. Danish, Norwegian and Swedish, will be discussed. The case studies are concerned with the modal auxiliaries ‘must/may’ (Chapter 3), the interrogative/epistemic adverb ‘I wonder’, the epistemic adverb ‘maybe’ (Chapter 4), and the discourse marker ‘I think’ (Chapter 5). These epistemic phenomena pose problems for traditional analyses within the framework of grammaticalization studies, because they are at the interface<sup>1</sup> of (most current definitions of) grammaticalization, lexicalization and pragmaticalization.

In order to clarify these phenomena, I will review various definitions and descriptions of lexicalization, grammaticalization and pragmaticalization, as well as the prototypical properties of these types of language change. The notions of lexicalization, grammaticalization and pragmaticalization will be redefined in such a way that they become complementary to each other. The converging and diverging properties of these different types of language change will be compared in order to determine the unique properties of lexicalization, grammaticalization and pragmaticalization, as well as their shared properties.

As regards subjectification, a specific type of semantic change involved in the development of epistemic modality, the three main perspectives (Langacker, Traugott and Nuyts) will be discussed and contrasted. In the present study, (inter)subjectification may affect linguistic

---

<sup>1</sup> In grammaticalization studies, ‘interface’ is a generally accepted term to denote overlap areas. It differs considerably from the generative use of this notion. In generative terms ‘interface’ relates to ‘interface rules’ and the ‘Interface Level.’ Interface rules regulate the mapping of the different components of grammar (a syntactic structure consisting of features, a morphological (and eventually phonological) structure, and a semantic interpretation). The interface rules apply at the Interface Level, commonly known as Logical Form (LF), which is usually taken to be the terminal syntactic object (cf. Adger 2003:31-2, 145).

items at different linguistic layers (ideational, textual and interpersonal level, cf. Traugott's tendencies in semantic change). It may, but need not, accompany different types of language change.

Ultimately, I will propose to give up the idea that linguistic changes can be assigned to predefined categories which are defined in terms of a specific set of semantic and/or formal properties. Instead, I will argue that it is more sensible to reduce lexicalization, grammaticalization and pragmaticalization to their basic mechanisms of change, primitive changes at the level of phonology/phonetics, morphology, syntax, semantics and discourse/pragmatics (Norde 2009:36), and the side effects or concomitants of change that may be used to identify potential instances of lexicalization, grammaticalization or pragmaticalization.

It will be shown that primitive changes tend to form clusters in terms of reduction or expansion<sup>2</sup> (e.g. semantic bleaching versus enrichment, morphological fusion versus separation) or no change at all. These clusters may correspond to changes traditionally labeled 'grammaticalization', 'lexicalization' or 'pragmaticalization', but changes may also cluster in alternative ways. This is what happens in the case studies in the present thesis. The advantage of a clustering approach is that it is capable of dealing with borderline cases and interfaces between different types of language change.

This chapter is organized as follows: in Section 1.1 the concept of epistemic modality is discussed and defined within the broader domain of modality. Section 1.2 is an introduction to the Scandinavian languages, their interrelations and history. The framework of grammaticalization studies is elaborated on and contrasted with formal approaches to language change in Section 1.3. Finally, Section 1.4 describes the method and sources, the aims and scope of the study, and provides an overview of the organization of this thesis.

## 1.1 Modality

Modality is a superordinate term which, very broadly defined, refers to any kind of speaker modification of an utterance. According to Nuyts (2006:1) modality is best described as a more abstract supercategory that consists of a set of more specific semantic categories, as cited in (2).

- (2) [T]he domain is usually characterized by referring to a set of more specific notions, each of which is defined separately, and which may be taken to share certain features motivating their grouping together under the label *modality*, but which differ in many other respects. As such, the notion of modality is best viewed as a supercategory (Nuyts 2005), which is much more loosely structured -and in fact probably belongs at a higher level of abstraction- than categories such as time and (types of) aspect.

---

<sup>2</sup> See also Traugott (2010a) who identifies two major approaches to grammaticalization, namely i) grammaticalization as reduction and ii) grammaticalization as expansion.

Depraetere & Reed (2006:269, my emphasis) list various semantic notions that are included in the domain of modality as well as properties that are shared by all modal utterances, as cited in (3) below.

- (3) The term ‘modality’ is a cover term for a range of semantic notions such as **ability**, **possibility**, **hypotheticality**, **obligation**, and **imperative meanings**. This is a serviceable definition for practical purposes. If, however, we wish to provide a more theoretically useful definition, we need to find what it is that all modal utterances have in common. [...] One feature that is common to all modal utterances is that they do not represent situations as straightforward facts [...]. We can get nearer to a positive characterization of modality if we say that modal meaning crucially involves **the notions of necessity and possibility**, or, rather involves a **speaker’s judgment** that a proposition is possibly or necessarily true or that the actualization of a situation is necessary or possible.[W]e shall work on the basis that all modal utterances are **non-factual**, in that they do not assert that the situations they describe are facts, and all involve the speaker’s comment on the necessity or possibility of the truth of a proposition or the actualization of a situation.

Hence, non-factivity, speaker’s judgment and necessity and possibility are the key notions in modality.

As is evident from the definitions in (2) and (3) above, modality is a complex semantic domain that consists of several subtypes. Numerous divisions have been proposed but there is no consensus on the number and types of modal subdomains. Perkins (1983:10) notes that “the number of modalities one decides upon is to some extent a matter of different ways of slicing the same cake.” Thus, there is no generally accepted way of dividing the modal domain, but it is important that one makes clear how one uses terms and definitions, as emphasized by van der Auwera & Plungian (1998:80).

To reorganize or redefine the many subtypes of modality into a conveniently arranged overview goes beyond the scope of the current investigation. Therefore, only epistemic modality will be thoroughly discussed in Section 1.1.3. The closely related domains of dynamic modality (Section 1.1.1) and deontic modality (Section 1.1.2), as well as the grammatical category of evidentiality (Section 1.1.4), will be examined because these four semantic domains are related to epistemic modality, and as such relevant to the present study. Section 1.1.5 provides a brief summary of the interrelations between epistemic modality and related domains.

### 1.1.1 Dynamic modality

Dynamic modality is concerned with physical or mental ability/capacity ‘be able to/know how to’ and potential. It also includes root (=general) possibility. Root possibility differs from epistemic possibility in that it is about general possibilities rather than estimations of likelihoods on behalf of the speaker. Examples of ability/capacity and general possibility meanings are given in (4) below.

- (4) a. *She can speak Spanish.*
- b. *I can swim.*
- c. *You can take bus 4 to get to the station (but also bus 1 or 8).*

Dynamic modality may give rise to deontic modality (Section 1.1.2) and epistemic modality (Section 1.1.3). The diachronic relations between these different modal domains will be discussed in more detail in Chapter 3.

### 1.1.2 Deontic modality

Deontic modality is traditionally defined in terms of permission and obligation as exemplified in (5a,b) respectively.

- (5) a. *John may come in now.*
- b. *John must come in now.*

It has been argued that a comprehensive definition of deontic modality should capture a degree of moral desirability of the state of affairs expressed in the utterance (Nuyts 2006:4-5). That is, widely defined, deontic modality may also relate to societal norms or personal (ethical) criteria, as exemplified in (6).

- (6) *John should apologize for what he has done.*

Because the focus of this study is on epistemic modality, the domain of deontic modality is broadly defined, as there is no need for a fine-grained taxonomy of deontic subtypes. The present use covers many subcategories and includes meanings associated with necessity, obligation, permission and (moral) desirability. Notions of (degrees of) moral acceptability or necessity are also included in the domain of deontic modality (cf. Nuyts, Byloo & Diepeveen 2010).

### 1.1.3 Epistemic modality

Epistemic modality, i.e. a speaker's evaluation of the likelihood of a state of affairs as conveyed in language (Nuyts 2001:xv), can be expressed in various ways (Nuyts 2001:29): by modal sentence adverbs as in (7a), predicative modal adjectives as in (7b), mental state predicates (modal lexical verbs, modal tags or parentheticals) as in (7c), or modal auxiliaries or modals as in (7d). These categories constitute the four major epistemic expression types.

- (7) a. *Maybe/probably/certainly/...they have run out of fuel*
- b. *It is possible/probable/likely/certain/...that they have run out of fuel*

- c. *I think/believe/...they have run out of fuel*
- d. *They may/might/must/...have run out of fuel*

Other ways of expressing epistemic modality are epistemic nouns (*possibility, probability, chance* etc.), verbal inflections (e.g. the subjunctive in German, French and Spanish), epistemic modal particles (e.g. in German, Dutch and Mainland Scandinavian), modal idioms, modal affixes, modal case, certain tenses (e.g. modal use of future or past tense) and prosodic features of intonation and stress (i.e. “the speaker’s doubt or uncertainty is regularly conveyed by a fall-rise intonation pattern”, Hoyer 1997:3). See de Haan (2006:32-42), Nuyts (2001:29) and Hoyer (1997:1-7) for examples and a more thorough discussion of these less central expression types.

According to Nuyts (2006:6), the core definition of epistemic modality is relatively clear, as “it concerns an indication of the estimation, typically, but not necessarily, by the speaker, of the chances that the state of affairs expressed in the clause applies in the world.” However, the literature on epistemic modality is far from consistent when it comes to a definition of this semantic domain. Some descriptions and definitions of epistemic modality from the functionalist literature<sup>3</sup> are provided in (8) below.

- (8) a. Palmer (2001:8): with epistemic modality speakers express their judgments about the factual status of the proposition.

Palmer distinguishes three kinds of “epistemic judgments” that according to him may be regarded as typological categories: Speculative, Deductive and Assumptive, as exemplified below (ibid:6).

Speculative:	<i>Kate may be at home now</i>	[a possible conclusion]
Deductive:	<i>Kate must be at home now</i>	[the only possible conclusion]
Assumptive:	<i>Kate will be at home now</i>	[a reasonable conclusion]

- b. Bybee et al. (1994:179-80): Epistemic modality applies to assertions and indicates the extent to which the speaker is committed to the truth of the proposition. The unmarked case in this domain is total commitment to the truth of the proposition, and markers of epistemic modality indicate something less than a total commitment by the speaker to the truth of the proposition. The commonly expressed epistemic modalities are possibility, probability, and inferred certainty.

---

<sup>3</sup> Formal approaches, such as modal logic and formal semantics, define (epistemic) modality in terms of ‘operators’ and ‘possible worlds’. Modal notions are represented by the sentential operators  $\Box$  (necessity) and  $\Diamond$  (possibility). For example,  $\Diamond p$  represents the statement *p is possibly true*. The notion of ‘possible world’ is a methodological tool based on the assumption that ‘worlds other than ours’ exist in a theoretical sense. Possible worlds are abstract entities that are used to model semantic relations among linguistic expressions (Kaufman, Condoravdi & Harizanov 2006:76).



- c. Coates (1983:18): [Epistemic modality] is concerned with the speaker's assumptions or assessment of possibilities and, in most cases, it indicates the speaker's confidence (or lack of confidence) in the truth of the proposition expressed.
- d. Perkins (1983:10): The term 'epistemic' derives from *episteme*, the Greek word for knowledge; however, the key concept which underlies modality seems to me to be the state of lack of knowledge which has been referred to by linguists in terms of 'non-factivity' (cf. Kiparsky and Kiparsky, 1970). To know (KNOW is a factive predicate) that a proposition is true presupposes that it actually is true; whereas, say, to be certain (CERTAIN is a non-factive predicate) that a proposition is true does not presuppose that it is true. It is true, of course, that one can 'know' something and be mistaken, but then one's 'knowledge' is, by implication, no longer knowledge. It must, therefore, be 'knowledge', as opposed to knowledge, which is relevant to a discussion of epistemic modality as it is understood here.
- e. Hoyer (1997:42-3): Epistemic modality is concerned with matters of knowledge or belief on which basis speakers express their judgments about states of affairs, events or actions. In the epistemic interpretations of (1) and (2), the speaker is clearly not making statements of fact or categorical assertions but conveying his subjective view of the world.
  - (1) *You may have a car*
  - (2) *You must be very patient*

Epistemic modality is a much-studied phenomenon both in linguistics and philosophy. These two disciplines are closely intertwined and the linguistic literature on epistemic modality is heavily influenced by the philosophical branch of modal logic. Some of the definitions cited above focus on 'the truth of the proposition' whereas others are concerned with 'matters of knowledge.' These different foci represent the 'difference' between alethic and epistemic modality, which is primarily a theoretical one.

Alethic modality (modes of truth) and epistemic modality (modes of knowing) differ in the sense that the former concerns the necessary, possible, contingent and impossible truth of a proposition, the latter pertains to the verified (known to be true), falsified (known to be false) and undecided (neither known to be true nor known to be false) truth of a proposition (von Wright 1951:1-2). In practice, however, alethic and epistemic modalities cannot be distinguished, as noted by Palmer (1986:11), because there is no difference between "what is logically true and what the speaker believes, as a matter of fact, to be true." He therefore concludes that "there is no formal grammatical distinction in English, and, perhaps, in no other language either, between alethic and epistemic modality."

Despite the large amount of research, there is still no consensus on how to define epistemic modality and how to demarcate it from other (modal) domains. The most common distinction within the field of modality is between root modality (also known as ‘non-epistemic modality’) and epistemic modality (also termed ‘non-root modality’). Root modality comprises both dynamic modality and deontic modality. Epistemic modality is occasionally distinguished from evidentiality, but in other cases evidentiality is subsumed under epistemic modality. The related domain of evidentiality will be further discussed in Section 1.1.4 below.

Frequently mentioned dichotomies in the description of epistemic modality are the splits between subjective and objective epistemic modality and between epistemic necessity and epistemic possibility. The distinction between objective modality (quantifiable possibility) and subjective modality (non-quantifiable possibility)<sup>4</sup> was introduced by Lyons (1977:800), who defines it in the following way in (9).

- (9) Subjective epistemic modality can be accounted for [...] in terms of the speaker’s qualification of the I-say-so component of his utterance. Objectively modalized utterances (whether their modality is alethic or epistemic) can be described as having an unqualified I-say-so component, but an it-is-so component that is qualified with respect to a certain degree of probability, which, if quantifiable, ranges between 1 and 0. If the factuality of an epistemically modalized proposition (as it is presented by the speaker) is of degree 1 it is epistemically necessary; if its factuality is of degree 0 it is epistemically impossible.

The division into objective and subjective epistemic modality is still very common in recent linguistic literature on epistemic modality. Necessity and possibility are the central notions (or modal operators) of traditional modal logic and are related to each other in terms of negation (represented by  $\sim$ ), as represented by example (10).

- (10) a. It is necessary that  $p$   $\equiv$  It is  $\sim$ possible that  $\sim p$   
b. It is possible that  $p$   $\equiv$  It is  $\sim$ necessary that  $\sim p$

That is, “if  $p$  is necessarily true, then its negation,  $\sim p$ , cannot possibly be true; and if  $p$  is possibly true, then its negation is not necessarily true” (Lyons 1977:787). This difference can be exemplified with deductive MUST and speculative MAY, as in (11a,b) respectively, which corresponds to the examples in (10a,b).

- (11) a. *Kate **must** be at home now* [epistemic necessity]

---

<sup>4</sup> For example, the proposition *Alfred may be unmarried* can be quantified for possibility in a situation like this: there is a community of ninety people, Alfred is one of them, thirty of them are unmarried, and it is unknown who is married and who is not. The possibility of Alfred being unmarried is presentable as an objective fact, there is a probability of (1/3) that the proposition *Alfred is not married* is true (Lyons 1977:798).

- b.      *Kate **may** be at home now*      [epistemic possibility]

Epistemic modality can also be viewed as a continuum instead of a binary notion that consists of possibility and necessity as distinct categories. On the continuum view, the semantic domain of epistemic modality is basically concerned with three concepts in both a positive and negative sense, namely (un)certainty, (im)possibility and (im)probability.<sup>5</sup> These can be placed along a so-called ‘epistemic scale’ that ranges from absolute certainty to absolute impossibility. The term ‘certainty’ is at the extreme positive end of the scale, ‘probability’ is situated around the middle of the positive side of the scale, and ‘possibility’ is near or at the neutral point on the positive scale, in the middle of the positive and negative side of the scale. This is, however, not an absolute ranking of the basic concepts within the domain of epistemic modality. The context in which they occur can modify the degree of likelihood that these concepts express (Nuyts 2001:55).

The negative side of the epistemic scale is thoroughly discussed in Ramat & Ricca (1998). They have split up the epistemic continuum in a positive and negative epistemic scale. These scales coincide with the positive and negative quantifier scale, respectively (Ramat & Ricca 1998:226), as shown in (12).

- (12) a.      certain            >      probable            >      possible  
              all                 >      many               >      some
- b.      impossible       >      improbable       >      uncertain  
              none             >      few                >      not all

These scales represent an epistemic adjectival continuum. Except for epistemic nouns and adjectives, a full epistemic scale (both a positive and negative side) is not applicable to modal adverbs, auxiliaries and mental state predicates, see Table 1 below.

		0%	-	← epistemic scale →		+	100%
	ADJ	impossible	improbable	uncertain	possible	probable	certain
	N	impossibility	improbability	uncertainty	possibility	probability	certainty
	ADV	*impossibly	*improbably	*uncertainly	possibly	probably	certainly
	AUX	-	-	may/can		must/will	-
	VP	-	-	I doubt	I believe	I think	I know

**Table 1.** Epistemic scale for adjectives, nouns, adverbs, auxiliaries and mental state predicates.

The present study adheres to a gradient view of epistemic modality in which necessity and possibility are not strictly binary notions because there are many modal meanings at the interface

<sup>5</sup> Speakers can further scalarize the basic concepts of epistemic modality by means of grading expressions (*very probable, rather certain, not entirely impossible*, etc.) or even quantify likelihood (e.g. *a 90 percent chance*), see Nuyts (2001:22).

of these apparently distinct domains (see Chapter 3). Moreover, degrees of possibility and necessity can be modified by means of various grading expressions (e.g. *very*, *rather*, *not* etc.). The working definition in this study, is the one formulated by Nuyts (2000:103), as quoted in (13) below.

- (13) By epistemic modality I mean the speaker's (expression of an) evaluation of the chances that the state of affairs talked about does or does not occur in the world.<sup>6</sup> This evaluation can range on an (epistemic) scale going from certainty that the state of affairs applies, via a neutral or agnostic stance, to certainty that it does not apply, with intermediary stages on the positive and negative sides of the scale ((im)probability).

The rise of epistemic meaning is a cross-linguistically regular type of semantic change which figures prominently in grammaticalization studies. In this study, no distinction is made between objective and subjective modality (e.g. Lyons in (9)) because likelihood estimations on behalf of the speaker are always subjective. As such, the development of epistemic meanings always involves subjectification, i.e. "[...] the development of a grammatically identifiable expression of speaker belief or speaker attitude to what is said" (Traugott (1995:32). The concept of subjectification will be thoroughly discussed in Section 2.1.4.

#### 1.1.4 Evidentiality

Evidentiality is concerned with the source of information on which an utterance is based. It states how the speaker/writer got to know the information s/he is presenting. For example, whether the 'evidence' is sensory or experiential (e.g. visual/auditory), first or second hand news (e.g. hearsay/reported speech), or inferred/reasoned from general knowledge or other available information. Some examples are given in (14) below.

- (14) a. *I see John getting on the bus.*  
 b. *I have been told that John did not come home last night.*  
 c. *John seems (to be) tired.*

Most linguists consider evidentiality, also termed 'evidential modality', to be a (sub)type of epistemic modality (e.g. Palmer 1986 [2001]), however, it has also been claimed that evidentiality is a category in its own right (e.g. Aikhenvald 2004, Nuyts 2001). The apparent resemblance between epistemic modality and evidentiality arises from some widespread misunderstandings and generalizations about the nature and interrelations between epistemic modality and evidentiality. The semantic difference between epistemic modality and

---

<sup>6</sup> Note that a speaker can also report on someone else's evaluation (no speaker commitment), e.g. *John thinks they have run out of fuel*, this is called descriptive use (Nuyts 2001: 39-41). See also Chapter 5 on 'I think.'

evidentiality is summarized as follows by de Haan (1999), in (15a,b), and Aikhenvald (2004:392), in (15c,d).

- (15) a. [Epistemic modality] **evaluate[s]** the speaker's commitment for the statement.
- b. [Evidentiality] **assert[s]** the nature of the evidence for the information in the sentence.
- c. Epistemic modality is modality concerned with epistemic meanings, i.e., meanings of (a) possibility or probability of an event or (b) of the reliability of information.
- d. Evidentiality is the grammatical marking of information source.

In Germanic and other European languages evidentiality is formally expressed through adverbs such as *apparently* and *evidently* and verbs like *seem*, *hear* or *see* (as in example (14)). But in other languages, such as Tariana (a North Arawak language spoken in northwest Amazonia), evidentiality is a grammatical category. That is, every utterance must contain a specification of the type of evidence on which it is based, as shown in (16) where *-pidaka* marks reported evidentiality, example from Aikhenvald (2004:3).

- (16) *Juse irida di-manika-pidaka*  
 José football 3sgnf-play-REC.P.REP7  
 'José has played football (we were told)'

Hence, every language has some way of referring to the source information, but not every language has a grammatical category of evidentiality.

In the present study epistemic modality and evidentiality are seen as two distinct, but related, domains. Stating the source of information is not the same as evaluating the likelihood of a state of affairs. Epistemic modality and evidentiality intersect when it comes to 'inference', i.e. epistemic necessity and inferential evidentiality constitute an overlap area (van der Auwera & Plungian 1998:86).

### 1.1.5 Epistemic modality and related domains

In sum, there are many ways of dividing the modal space. There is, however, no consensus on the terminology which is reflected by a wide variety of terms that denote the same concept. For an overview see Nuyts 2006; van der Auwera & Plungian 1998; Palmer 2001. A rough

<sup>7</sup> The abbreviations stand for: 3=third person, sg=singular, nf=non-feminine, REC=reciprocal, P=past, and REP=reported.

distinction in the modal domain is between epistemic and root (= dynamic + deontic) modality. In the present study a narrow definition of modality will be applied, in the sense that the modalities in the current investigation can be defined in terms of (degrees of) necessity and possibility. That is, the expressions under investigation in the present study, i.e. epistemic adverbs, modal auxiliaries and mental state predicates, convey meanings in the realm of necessity and possibility, as well as meanings that arose out of these contexts (i.e. various speech-act meanings (Chapter 4 and 5) and so-called postmodal meanings (Chapter 3)). Evidentiality is taken to be a related, but distinct category, which overlaps when it comes to inferred or reasoned meanings.

## 1.2 The Mainland Scandinavian languages

The Scandinavian language area consists of five official Nordic languages with Germanic roots (Danish, Norwegian<sup>8</sup>, Swedish, Faroese and Icelandic). In addition to these languages, Finnish, Greenlandic, Saami and various immigrant languages are spoken in present-day Scandinavia.

From a historical perspective, four different divisions can be made for the Scandinavian languages. These are illustrated in Table 2-4 and based (=translated and slightly adapted) on Torp (1998:34-35). The more lines are drawn between the different varieties, the more they have diverged from one another. The first division is between East and West Nordic. Danish and Swedish belong to the East Nordic subgroup whereas Norwegian forms the West Nordic subgroup. This division reflects the language situation during the Viking age, as represented by Table 2 below.

<b>West Nordic</b>	<b>East Nordic</b>
(Norwegian)	(Swedish and Danish)

**Table 2.** Language situation in Scandinavia during the Viking age (± 800).

A second stage of language division arose around the year 1200. Norwegian (together with Faroese and Icelandic) and Swedish form the North Nordic subgroup whereas Danish forms the South Nordic subgroup on its own, as illustrated by Table 3.

<b>North Nordic</b>		<b>South Nordic</b>
Norwegian (Faroese, Icelandic)	Swedish	(=Danish)

**Table 3.** Language situation in Scandinavia around the year 1200.

<sup>8</sup> Norwegian has two official written varieties *bokmål* 'book language' and *nynorsk* 'new Norwegian.' Bokmål developed out of Danish and Dano-Norwegian (between 1380–1814, Norway was part of the Kingdom of Denmark, between 1814–1905 Norway was part of the Kingdom of Sweden. Norway is an independent country since 1905), and is influenced by Norwegian vernacular. Nynorsk is 'composed' on the basis of a variety of Norwegian dialects. Its foundations were laid by Ivar Aasen (1813–1896). In the present study only *bokmål* is taken into account.

The language situation around the year 1500 resembles the classification of Scandinavian languages that is still used these days. Within the modern Scandinavian languages there are two main groups: Danish, Norwegian and Swedish form the Mainland Scandinavian language group and Icelandic and Faroese form the Insular Scandinavian language group, see Table 4.

Insular Nordic				Mainland Scandinavian			
Icelandic			Faroese			North Scandinavian	South Scandinavian
						Norwegian	Swedish
							Danish

**Table 4.** Language situation in Scandinavia after the year 1500.

Danish, Swedish and Norwegian have lost most of the original Old Norse case system and are strongly influenced by Middle Low German and international loan words. Icelandic and Faroese on the other hand, have preserved an extensive case system and remained relatively unaffected by Middle Low German and other loan words. The present-day language situation in Scandinavia can be illustrated as in Table 5.

Insular Nordic		Mainland Scandinavian	
Icelandic		North Scandinavian	
		South Scandinavian	
		Nynorsk	
	Faroese	Swedish	Bokmål
			Danish

**Table 5.** Language situation in present-day Scandinavia.

In general, Danish, Norwegian and Swedish are quite similar when it comes to vocabulary, sound system, inflection and grammar. Danish and Norwegian are most similar with regard to vocabulary (both function and content words) whereas the sound system is more similar for Norwegian and Swedish. Both the sound system and vocabulary are more divergent for Danish and Swedish. See Braunmüller (1998); Torp (1998) for a comprehensive overview of the history and differences and similarities for the Scandinavian languages.

The Mainland Scandinavian language area is a dialect continuum, i.e., an area in which neighbouring dialects are mutually intelligible. In Scandinavia, the neighbouring dialects change gradually without sharp borders until Saami borders off the continuum in the north, Finnish in the east and German in the south. The standard varieties of the Mainland Scandinavian languages are to a large degree mutually intelligible. That is, speakers of these languages can communicate in their own language without the interference of a lingua franca. Icelandic and Faeroese (Insular Scandinavian) are still related languages but semi-communication is not possible due to larger linguistic distances (e.g. Gooskens et al. 2008).

Both synchronically and diachronically, epistemic modality has been studied in great detail for some languages (for Dutch and German see e.g. Nuyts 2001; for English see e.g. Traugott 1989, 1995). As regards the Mainland Scandinavian languages no detailed synchronic,

diachronic and comparative study of epistemic phenomena has been carried out so far, except for short descriptions in (historical) reference grammars and specialized case studies (dissertations), e.g. Andersson (2007). As pointed out by de Haan & Hansen (2009:2) “there is a considerable lack of work comparing the degree of the grammaticalisation of modal elements in different languages”, a topic which is thoroughly addressed in their book on modals in the languages of Europe (Hansen & de Haan 2009).

Since the Mainland Scandinavian languages are Germanic languages too, one might expect the epistemic expressions in these languages to develop according to the observed tendencies for its cognates in the Germanic languages. Being closely related language varieties, the Mainland Scandinavian languages can be expected to have gone through similar developmental stages and have comparable degrees of grammaticalization for their cognate epistemic expressions. Hence, the Mainland Scandinavian languages are both suitable and interesting for the present study from a descriptive, comparative and theoretical point of view.

### **1.3 The framework of grammaticalization studies**

The study of grammaticalization and related phenomena is a flourishing branch of historical linguistics. The main question within grammaticalization studies is ‘how does grammar arise?’ or more specifically ‘how do grammatical elements come into being?’ Though the search for the origins and development of grammatical categories of language is not new, grammaticalization phenomena have received a lot of attention, especially in the last decades of the 20<sup>th</sup> century. This is reflected by a wide variety of descriptive and theoretical studies devoted to grammaticalization phenomena. Several collected works and volumes (Heine et al. 1991; Traugott & Heine 1991; Hopper & Traugott 2003; Narrog & Heine 2011), journals (e.g. special issues of *Language Sciences*, Campbell 2001; Norde et al. forthc.), and publications on the occasion of conferences and workshops with grammaticalization as the central theme, e.g. *New Reflections on Grammaticalization* (Wischer & Diewald 2002; Fischer et al. 2004), have been published since the 1990’s. For an overview of the history of grammaticalization (studies), see Hopper & Traugott (2003:19-38), Lehmann (1995:1-8), Heine et al. (1991:11-220).

In its early days, the study of grammaticalization (phenomena) used to be presented and formulated as a theory. It was seen as a specific type of change in its own right centered on the principle of unidirectionality and characterized by different aspects of (semantic) change. It could also be exploited as a tool for historical reconstruction. This view on grammaticalization has led to serious criticism (cf. especially the contributions by Joseph, Janda and Newmeyer to a special issue of *Language Sciences* entitled ‘Grammaticalization: a critical assessment’). Especially proponents of generative linguistics have questioned the basic tenets of so-called ‘grammaticalization theory.’ For example, Newmeyer (2001:188) concludes that “there is no such thing as grammaticalization, at least in so far as it might be regarded as a distinct grammatical phenomenon requiring a distinct set of principles for its explanation.” This is true insofar that grammaticalization is a composite change that consists of a series of changes that are not specific to grammaticalization (see Chapter 2). However, the correlation of several changes



that collectively form a composite change makes grammaticalization stand out as a distinct type of language change.

The following citation by Campbell (2001:113), in (17), nicely captures the generally held view on grammaticalization these days.

- (17) This [i.e. grammaticalization is not a distinct grammatical phenomenon requiring a distinct set of principles for its explanation. KB] raises the question of whether grammaticalization has any value at all. The position taken here is that it does, that while the phenomenon of grammaticalization is interesting and attention on it has provided a range of valuable information, there are serious problems with so-called “grammaticalization theory.”

It is now more common to see grammaticalization as an empirical phenomenon with descriptive value, not a theory with explanatory power. Grammaticalization is a useful concept in describing language change and variation. In this study the term ‘grammaticalization studies’ is used to refer to the study and framework of grammaticalization as an empirical phenomenon.

### **1.3.1 Approaches to grammaticalization**

Within the framework of grammaticalization studies there are different approaches, perspectives and objectives. Studies may be descriptive or theoretical, undertaken from a synchronic or diachronic perspective, centered on the whole process or on the outcome of the process, or focus primarily on formal change, semantic change or ‘change in progress.’

As for linguistics in general, there are both functional and formal approaches to grammaticalization. These approaches are fundamentally different and the divergent views have led to heated discussions which have not been settled until today. The basic assumptions of formal and functional approaches with regard to, *inter alia*, variation, change and categorization are largely incompatible with each other.

However, both approaches offer valuable insights into language change, which makes them, despite their insurmountable differences, complementary to each other. See Fischer & Rosenbach (2000:8-14) for an overview and comparison of formal and functional approaches to grammaticalization.

### **1.3.2 Formal approaches to grammaticalization**

Grammaticalization, and language change and variation in general, are problematic issues for formal or generative models of grammar. In this line of thought, language change is abrupt (=reanalysis), complete (=only discrete word class categories) and language internal (=i-language, i.e. language as an innate capacity). Language change takes place from one generation to the next during the process of language acquisition. Change is manifested either as the evolution of functional categories or as altered movement operations or changes in functional categories.

Variation can only be accounted for within the so-called ‘double-base hypothesis’, i.e. an individual may have access to more than one grammar simultaneously (Pintzuk 1991). There is a strict separation of synchrony and diachrony in that change is a comparison of the speaker’s competence before and after the change. Semantic-pragmatic factors, as well as underlying motivations for the change, are not taken into account because the focus is on syntax as an autonomous system (cf. ‘the hypothesis of autonomous syntax’, Chomsky 1957:17).

Grammaticalization in generative terms, then, is concerned with syntactic change only. It is “an instance of upwards reanalysis, which gives rise to new functional material” (Roberts & Roussou 2003:205). Hence, in formal terms grammaticalization is a process of syntactic simplification by which elements are merged higher up in the syntactic structure and lexical heads are reanalyzed as functional heads. See, amongst others, the work of Lightfoot (1979), Roberts (2007), van Gelderen (2004) for diachronic studies within the generative framework.

### **1.3.3 Functional approaches to grammaticalization**

Most studies in grammaticalization are carried out within the functional linguistic tradition. In functional linguistics, language change is seen as a gradual process that is shaped by linguistic and extra-linguistic factors. Diachrony is considered to be an integral part of synchrony because the whole process of change, from the beginning to the end, is taken into account.

The focus is on all facets of change, i.e. changes in syntax, semantics, morphology and the like. Change takes considerable time to develop. Forms and structures are not abruptly replaced but they gradually acquire properties of another category or construction. Original and newly developed forms may co-exist for centuries. As such variation is a logical consequence of the gradient nature of categories and the gradualness of language change in general.

Recent trends in grammaticalization studies include, amongst others, the relation between grammaticalization, (inter)subjectification and pragmaticalization (e.g. Degand & Simon-Vandenberg 2011), degrammaticalization and exaptation (e.g. Norde 2009), and the incorporation of grammaticalization into other linguistic frameworks, such as generative grammar (Roberts & Roussou 2003; Roberts 2007; van Gelderen 2004) or various types of (diachronic) construction grammar (e.g. Traugott & Trousdale 2010; Trousdale *forthc.*).

### **1.3.4 The approach to grammaticalization in the present study**

In the present study the term ‘grammaticalization studies’ covers (the study of) grammaticalization and related concepts, such as lexicalization, pragmaticalization and (inter)subjectification that arose in the context of grammaticalization studies. A functional approach to grammaticalization is taken because of the focus on language use, diachrony in synchrony, the gradualness of language change and the incorporation of conceptual, semantic-pragmatic and language external factors that may be involved in language change. Semantic and formal changes are considered to be equally important.

Grammaticalization is seen as an empirical phenomenon with descriptive value, not a theory with predictive and explanatory power. On the basis of identified developmental

tendencies one can make hypotheses about change, but one has to keep in mind that grammaticalization is by no means deterministic. Change occurs through language use and may be shaped by extra-linguistic factors, hence detailed diachronic studies are always needed to verify hypothesized developmental paths. Grammaticalization studies provides diagnostics to identify tendencies in language change and is an accurate framework to describe language change and variation in one, or several, language(s).

## 1.4 The present study

The present study is a comparative corpus investigation of epistemic expressions in the Mainland Scandinavian languages and consists of a descriptive and a theoretical part. These are closely intertwined as corpus data is used to clarify, illustrate and explain theoretical issues. The study is primarily synchronic in nature. It concerns an empirical (=data-based) and contrastive investigation on the basis of modern online corpora and reference grammars, which is occasionally supplemented with historical data from earlier stages of these languages.

The core of this thesis is formed by three case studies from the domain of epistemic modality, i.e. modal auxiliaries ‘must, may’ (Chapter 3), epistemic adverbs ‘maybe’ and ‘I wonder’ (Chapter 4) and mental state predicates of the type ‘I think’ (Chapter 5). Each of these epistemic expressions will be examined in a detailed case study in order to clarify theoretical issues by means of empirical data. The focus is on these expression types because they form the basic and most frequent means of expressing epistemic modality. Other expression types occur more marginally (e.g. nouns, particles, tense, and intonation) or are not or no longer applicable to the Mainland Scandinavian languages (e.g. the subjunctive).

### 1.4.1 Sources and method

The method and sources in this study are the same for all three case studies. The data have been selected from modern online corpora with the aim to analyze comparable samples, both in size and text type. Corpora provide actual instances of language use, information about the frequency with which an item occurs, as well as the larger contexts and collocations in which an item occurs. These are important factors in the study of ongoing change.

The Norwegian data have been collected from *Norsk Aviskorpus*.<sup>9</sup> This modern dynamic text corpus consists of newspaper texts from 1998 up to now. Texts from various types of newspapers are part of the corpus: tabloids, broadsheets, national and regional newspapers as well as general, business and financial newspapers. The corpus contains circa 640 million words and approximately 200.000 to 250.000 words of running text are being added to the corpus each day.

The Swedish data have been taken from *Språkbankens koncordanser*.<sup>10</sup> This corpus contains a subset of modern and historical online corpora and includes texts from various sources: newspapers, literature, non-specialist literature, government debates, law texts and historical texts. The entire corpus has approximately 99 million words. The subcorpora *Press95-*

---

<sup>9</sup> <http://avis.uib.no/>

<sup>10</sup> <http://spraakbanken.gu.se/konk/>

98, which all contain newspaper texts, provide the data for the Swedish samples.

The Danish data have been selected from *KorpusDK*<sup>11</sup>, which is a modern online corpus of Danish texts from different sources and genres, with a total of 56 million words. The texts were collected between the years 1990 and 2000. The corpus consists of two subcorpora: *Korpus 2000* which consists of texts from the period 1998 to 2002 and *Korpus 90* which consists of texts from the period 1983 to 1992. *KorpusDK* is mainly designed for linguistic analysis with the particular aim to identify recent developments in the Danish language.

The corpora *Norsk Aviskorpus*, *Språkbankens konkordanser Press95-98* and *KorpusDK* are comparable in the sense that they all contain newspaper texts, but *KorpusDK* contains literary texts as well, which makes it less comparable to the Norwegian and Swedish corpora.

A quantitative analysis based on the corpus data will be carried out. This method is particularly suitable to explore synchronic variation and establish prototypical and less central meanings/forms of a certain expression. Both occurrence (=positive evidence) and non-occurrence (=negative evidence) of certain syntactic/semantic features are relevant to the analysis. The counts per investigated variable (i.e. meaning, syntactic position etc.) are analyzed by means of a Chi-square test or Loglinear Analysis, depending on the number of variables involved in the analysis. Values for Cramer's V, a measurement to determine the strength of an association between categorical variables (Field 2005:689, 693), will also be given. More details on these statistic methods will be provided in the analysis of the corpus data in Chapter 3-5.

### 1.4.2 Aims and scope

The purpose of the present study is threefold. First, it aims to give an overview of the semantic and formal distributions, degrees of grammaticalization, lexicalization or pragmaticalization and the differences and similarities for the epistemic phenomena, as listed in (18a), in the Mainland Scandinavian languages. In particular, the following research questions of a general nature, as stated in (18), will be addressed.

- (18) a. What is the origin and the development of the epistemic expressions 'must, may', 'maybe', 'I wonder' and 'I think' in the Mainland Scandinavian languages?<sup>12</sup>

---

<sup>11</sup> <http://ordnet.dk/korpusdk>

<sup>12</sup> One might wonder how epistemic modality was expressed before the current system came into being. To this question I can only give a suggestive answer by observing that modal meanings were expressed through verbal means, such as mood, modal auxiliaries and verb phrases (*kann (svá) vera* 'may well be', cf. Fritzner 1867). Faarlund (2004:246) notes for Old Norse that "the indicative mood is used to express factivity, the subjunctive is used in a non-factive sense [and that t]he subjunctive has two different semantic functions, optative and potential." The auxiliary *munu* 'will, may' has epistemic meaning (intention or possibility) and *kunna* 'can' may function as an epistemic modal auxiliary and *skulu* 'shall' has deontic meaning (ibid:129). All verbal expressions may occur with indicative or subjunctive mood.

- b. What are the differences and similarities between Danish, Swedish and Norwegian with respect to the formal and semantic properties of these epistemic expressions? More specifically, do they differ in the degree of lexicalization, grammaticalization or pragmaticalization of these epistemic markers?

The second objective is to link the outcome of the case studies to more comprehensive theoretical issues that figure prominently in grammaticalization studies. The epistemic phenomena listed in (18a) will be examined in order to clarify some current controversies in theorizing about grammaticalization and related types of language change. In particular, the following research questions, as stated in (19), will be addressed.

- (19) a. What is the relation between grammaticalization and (inter)subjectification in the development of the Mainland Scandinavian modals ‘must,may’?
- b1. Is the development of Mainland Scandinavian ‘I wonder’, which is a shift from minor to minor category, a case of grammaticalization?
- b2. Is the development of Mainland Scandinavian ‘maybe’ a case of lexicalization, grammaticalization or both?
- c. Is the development of Mainland Scandinavian ‘I think’ a case of lexicalization, grammaticalization or pragmaticalization?

Third, as already noted, the epistemic phenomena in the present study defy strict categorization. The corpus data are discussed and analyzed with respect to characteristics of grammaticalization, lexicalization or pragmaticalization in order to find out which of these types of language change most appropriately captures their development.

Ultimately, I aim to overcome the problematic status of so-called ‘hybrid cases’ (i.e. items with properties of more than one type of language change) by means of a clustering approach. This approach is based on mechanisms of change, correlated primitive changes at different linguistic levels, and the concomitants of these (micro-)changes. It enables one to identify interface areas as well as prototypical instances of lexicalization, grammaticalization and pragmaticalization.

### **1.4.3 Organization of the study**

This dissertation consists of six chapters. This exploratory chapter, Chapter 1, introduced and motivated the topic of the current study. Chapter 2 is a preliminary chapter in which the concepts of grammaticalization, lexicalization, pragmaticalization, their interfaces and relation to (inter)subjectification will be discussed and defined.

Chapter 3 presents the first case study which is on modal auxiliaries. Modal auxiliaries are undoubtedly the most studied elements when it comes to epistemic modality. The present investigation concentrates mainly on (subtle) differences in meaning for the Mainland Scandinavian modal auxiliary ‘must, may’, degrees of grammaticalization and the role of (inter)subjectification in the development of (post)modal meanings.

Chapter 4 presents the second case study on modal adverbs, more specifically the interrogative/epistemic adverb ‘I wonder’ and the epistemic adverb ‘maybe.’ The development of epistemic adverbs of the type modal auxiliary ‘can’ or ‘may’ + a verb meaning ‘happen’ or ‘be’ is an excellent case to study the grammaticalization-lexicalization interface. This type of epistemic adverb is on the border between grammaticalization and lexicalization because it is not clearly a lexical or grammatical item. The development of the interrogative/epistemic adverb ‘I wonder’ is a shift from minor to minor status. The question is whether this is a change from grammatical to even more grammatical, and how one is to assess ‘more grammatical status.’ That is, does the development of ‘I wonder’ qualify as an instance of grammaticalization?

Chapter 5 is concerned with the third case study, which is about the mental state predicate ‘I think.’ Mental state predicates are very suitable to elaborate on another controversial topic within grammaticalization studies, namely the rise of discourse markers and the status of pragmaticalization. Pragmaticalization is a contentious issue in grammaticalization studies because it is not a generally accepted type of language. The rise of discourse markers adheres to many principles of grammaticalization, but their development also violates other fundamental characteristics of grammaticalization.

Finally, Chapter 6 presents a novel model of analysis on the basis of essential mechanisms of change, clusters of primitive changes and concomitants of these (micro-)changes, which collectively may identify a change as grammaticalization, lexicalization, pragmaticalization or a borderline case at the interface of these different types of language change. Moreover, Chapter 6 contains a discussion, the main conclusions of the present study, and an outlook to further research and future work.

## Chapter 2

### 2. Preliminaries

Lexicalization, grammaticalization, pragmaticalization and (inter)subjectification are all fuzzy notions that have been conceptualized in various ways. Lexicalization, grammaticalization and pragmaticalization are composite types of language change that consist of semantic reinterpretation and formal reanalysis, accompanied by a set of correlated primitive changes at different linguistic levels (phonology/phonetics, morphology, syntax, semantics and discourse/pragmatics) and the concomitants of these (micro-)changes. Subjectification and intersubjectification are specific types of semantic change that may, but need not, be involved in lexicalization and grammaticalization, whereas (inter)subjectification is a necessary component of pragmaticalization.

In order to compare lexicalization, grammaticalization and pragmaticalization in a meaningful way, they have to be defined in a similar manner. For reasons of consistency it is necessary to adapt and complement existing definitions in such a way that the notions of lexicalization, grammaticalization and pragmaticalization become comparable and complementary to each other. This is a prerequisite for a unified account of these related types of language change, which forms the basis for the description and analysis of the case studies on the development of the epistemic expressions ‘must, may’, ‘I wonder’, ‘maybe’ and ‘I think.’ Hence, it is not possible to use and compare existing definitions by different authors who might have different background assumptions or even contradicting views on lexicalization, grammaticalization and pragmaticalization.

This chapter is organized as follows: in Section 2.1 the notions of grammaticalization, lexicalization, pragmaticalization and (inter)subjectification will be discussed and defined. Subsection 2.2 deals with the interfaces between these different types of language change. The lexicalization-grammaticalization interface, grammaticalization-pragmaticalization interface, pragmaticalization-lexicalization interface, and the lexicalization-grammaticalization-pragmaticalization interface will be discussed on the basis of hybrid linguistic items that are situated at the interfaces of these different types of language change.

### 2.1 Definitions

#### 2.1.1 Grammaticalization

The term ‘grammaticalization’<sup>13</sup> is used in two different but related meanings. In the first place it denotes the theoretical framework of grammaticalization studies (Section 1.3), and secondly it refers to the gradual diachronic process of language change itself (Hopper & Traugott 2003:1-2). In other words, grammaticalization concerns the study of the origin and development of

---

<sup>13</sup> The notion of ‘grammaticalization’ is also known as ‘grammaticization’ (e.g. Hopper 1991; Bybee et al. 1991) or ‘grammatization’ (e.g. Matisoff 1991).

grammatical items or function words (also known as *grams*<sup>14</sup> in Bybee et al.'s (1994:2) terminology).

It is generally accepted that there is a distinction between lexical items (=content words) and grammatical items (=function words). Nouns (e.g. *car*), verbs (e.g. *write*), and adjectives (e.g. *blue*) are examples of lexical items. Lexical items are used "to report or describe things, actions, and qualities" (Hopper & Traugott 2003:4). Prepositions (e.g. *of*), connectives (e.g. *and*, *or*), pronouns (e.g. *it*), and demonstratives (e.g. *this*) are function words. Grammatical items "serve to indicate relationships of nominals to each other (prepositions), to link parts of a discourse (connectives), to indicate whether entities and participants in a discourse are already identified or not (pronouns and articles), and to show whether they are close to the speaker or hearer (demonstratives)" (ibid).

The basic idea in grammaticalization studies is that grammatical elements have their origin in lexical items, which is reflected in Meillet's (1912:131) famous and often-cited definition, as quoted in (20) below.

- (20) [Grammaticalization is KB] the attribution of a grammatical character to a previously autonomous word.<sup>15</sup>

The adjective 'autonomous' can be interpreted in different ways. In a formal sense it may refer to a so-called 'free morpheme', i.e. a morpheme that can stand on its own (e.g. nouns like *house*), as opposed to a 'bound morpheme' that is part of a larger linguistic unit (e.g. affixes like *-ly* in *possibly*). Semantically, it may denote a morpheme with referential meaning, i.e. a morpheme that has meaning of its own (e.g. nouns such as *table*), as opposed to morphemes with relational or indexical meaning whose meaning has to be interpreted in context (e.g. pronouns like *she*). The term 'grammatical character' is not further defined but is generally understood as 'grammatical function' or 'grammatical meaning.'

Not only lexical items, but also grammatical(ized) elements may be subject to (further) grammaticalization, as stated in Kuryłowicz's (1975 [1965]:52) classical definition of grammaticalization in (21).

- (21) Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a grammatical to a more grammatical status, e.g. from a derivative formant to an inflectional one.

This definition entails two components: a shift from lexical to grammatical status, and a shift from grammatical to even more grammatical status. It remains unclear how one is to assess

---

<sup>14</sup> The term 'gram' is a short form of 'grammatical morpheme', e.g. affixes, stem changes, reduplication, auxiliaries, particles or complex constructions like English *be going to* (Bybee et al. 1994:2).

<sup>15</sup> Translation of '*l'attribution du caractère grammaticale à un mot jadis autonome*' by Paul. J. Hopper (1991:131).



‘more grammatical status’ and on what grounds some grammatical expressions are more ‘grammatical’ than others. I will return to this issue in Chapter 3 and 4.

The shift from lexical to grammatical status is also known as ‘primary grammaticalization’, the shift from grammatical to even more grammatical status is called ‘secondary grammaticalization’. The distinction between primary and secondary grammaticalization is also found in Traugott (2002), and Hopper & Traugott’s (2003:18) definition of grammaticalization in (22).

- (22) [Grammaticalization is KB] the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions.

According to Detges & Waltreit (2002:188) only the first part of the definition, i.e. primary grammaticalization, counts as proper grammaticalization. Increasing grammaticalization, as denoted by the second part of the definition (i.e. secondary grammaticalization), is considered to be a subsequent stage of increasing bondedness or morphologization (e.g. cliticization, affixation).

What the definitions in (20), (21), (22) and many others have in common, is that they are primarily semantic definitions of grammaticalization. That is, these characterizations remain rather vague as to how ‘grammatical meaning’ or ‘grammatical status’ is formally expressed. The following descriptions by Heine & Reh (1984), in (23), and Fischer & Rosenbach (2000:2), in (24), give some more details about the micro-changes involved in grammaticalization.

- (23) With the term ‘grammaticalization’ we refer essentially to an evolution whereby linguistic units lose in semantic complexity, pragmatic significance, syntactic freedom, and phonetic substance, respectively (Heine & Reh 1984:15).
- (24) Grammaticalization is generally seen as a process whereby a lexical item, with full referential meaning (i.e. an open-class element), develops grammatical meaning (i.e. it becomes a closed-class element); this is accompanied by a reduction in or loss of phonetic substance, loss of syntactic independence and of lexical (referential) meaning (Fischer & Rosenbach 2000:2).

A more recent approach to grammaticalization focuses on the paradigmatic organization and unification of morphological, topological and constructional paradigms (Nørgård-Sørensen, Heltoft & Schøsler 2011), as described in (25).

- (25) [W]e offer an overall view of grammaticalization (grammatical status and grammatical change) radically different from the standard view(s) centering around what is called the cline of grammaticality. [...] Our emphasis is on the generalised concept of a paradigm,

its structuring role in grammar and its importance for a theory of linguistic change. [...] Morphological, topological and constructional paradigms very often connect to form complex paradigms, [...], and grammaticalisation processes include the formation, restructuring and dismantling of such complex paradigms (Nørgård-Sørensen, Heltoft & Schøsler 2011:i-xii).

Nørgård-Sørensen, Heltoft & Schøsler introduce the concept of ‘connecting grammaticalisation’ in which “chains of grammations, regrammations and degrammations are seen as one connected process (ibid:xiii). [...] The concept of connecting grammaticalization is intended to capture the fact that two or more simple paradigms can be related to each other in non-trivial ways and that such relations arise as historical formations (ibid:103).” This view seems to conflate grammaticalization with grammatical change, and as such to define grammaticalization in a very broad sense, so as to include word order changes and changes in (morphological) paradigms, developments which are traditionally excluded from definitions of grammaticalization.

Boye & Harder (2012) propose a usage-based theory of grammatical status and grammaticalization. They argue that grammatical expressions and grammaticalization cannot be defined in terms of specific phonological, morphosyntactic, or semantic features, alone or in combination, but that these notions “can be defined in terms of the ancillary status that grammatical expressions by linguistic convention have in relation to other expressions,” as cited in (26) and (27) respectively (ibid:7).

(26) Grammatical expressions are by convention ancillary and as such discursively secondary.

(27) Grammaticalization is the diachronic change that gives rise to linguistic expressions that are by convention ancillary and as such discursively secondary.

By ‘linguistic convention’ they mean the reflection of “a state of co-ordination between members of a community that goes beyond individual instances of linguistic communicative interaction (ibid:8). The notion of ‘ancillary status’ denotes that grammatical expressions (morphemes, words, constructions), in contrast to lexical expressions that may (or may not) convey the main point of a linguistic message, are conventionally specified as non-carriers of the main point, serving instead an ancillary communicative purpose as secondary background elements (ibid:6). Being ‘discursively secondary’ is “to have lower (discourse) prominence than one or more syntagmatically related expressions in the utterance” (ibid:8).

Boye & Harder distinguish between three basic types of grammaticalization (ibid:21-22), as listed in (28). The first type has its source in a lexical expression, the second type has its source in a ‘pragmatically conveyed’ meaning with a secondary status, and for the third type the source is an already existing grammatical expression.

- (28) a. [G]rammaticalization consists in ANCILLARIZATION, a CHANGE IN EXISTING DISCOURSE PROMINENCE CONVENTIONS: The potentially primary status of a lexical expression is replaced with the secondary status of a grammatical expression.
- b. [G]rammaticalization consists in a CONVENTIONALIZATION OF A DISCURSIVELY SECONDARY MEANING as a property of a new linguistic expression: A linguistic expression -for instance fixed word order- becomes conventionally associated with a secondary meaning which was originally part of a pragmatic total message, but not conventionally associated with any linguistic expression.
- c. [G]rammaticalization consists in the development of such an expression [i.e. an already existing grammatical expression KB] into a new grammatical expression distinct from its source. [...] Within this type a distinction can be made between three subtypes (ibid: 43):
- i) CONTINUED GRAMMATICALIZATION: the development of a grammatical expression into a new expression which is in some sense more grammatical than its source.
  - ii) DEGRAMMATICALIZATION: the development of a grammatical expression into a new expression which is less grammatical than its source.
  - iii) REGRAMMATICALIZATION: the development of a grammatical expression into a new expression which is neither more nor less grammatical than its source.

Their theory combines function and language use with linguistic structure, identifies differences and similarities between lexical and grammatical meaning, provides explicit tests to assess grammatical status (i.e. focus and addressability tests, ibid:14-15), and generalizes over controversial and non-controversial instances of grammaticalization.

So far, a couple of traditional and more recent definitions have been discussed. However, as noted by Narrog & Heine (2011) in *The Oxford handbook of grammaticalization*, which presents the state of the art in research on grammaticalization, there is still no generally accepted definition of grammaticalization:<sup>16</sup>

- (29) Going through the chapters of this volume, the reader will notice that grammaticalization is far from being a uniform concept, and various definitions have been proposed. [...] Still, when controversies arise many scholars agree in draw attention [sic.] to the classic

---

<sup>16</sup> See also Campbell & Janda (2001) for an extensive overview of definitions of grammaticalization, and Lindström (2004) for a thorough discussion of the history and meaning of the term and concept called grammaticalization.

definition by Kuryłowicz [see (21) above KB] to help settle the issue of what should be subsumed under the rubric of grammaticalization (Narrog & Heine 2011:2-3).

That this is ‘the-state-of-the-art’ in the year 2012 is emphasized once more by a two-day workshop<sup>17</sup> at the Free University of Berlin, entitled ‘So, what is it then, this grammaticalization?’ The aim of this meeting was refining the notion of grammaticalization, which itself had become bleached and eroded, “in a way that is beneficial for our understanding of language change.” The call for papers highlights one more terminological problem:

- (30) Most controversies concerning the properties and the status of grammaticalization have their origin in the fact that the notion has become inconsistent or even ill-defined. A further consequence is that a plethora of new *Izations* in the study of (grammatical) change have emerged, but no harmonious terminology – not to speak of a consistent model of the emergence and the change of grammatical forms (von Mengden & Simon 2012).

All in all, it seems impossible to come up with a generally accepted definition of grammaticalization<sup>18</sup> and related phenomena.

In the present study, grammar is the cover term for “the set of categories, patterns and organizing principles evidenced by language, most essentially abstract patterns of semantics, syntax, morphology, and phonology that at least in theory permit infinite combinations” (Brinton & Traugott 2005:4). Grammatical items are those elements that belong to a minor category, and that have relational meaning and secondary status. Their prime function is to regulate grammatical structure and grammatical relations. In the remainder of this section I will discuss the notion of grammaticalization in more detail and make clear how it is used and defined in this study

### **2.1.1.1 Grammaticalization as a (uni)directional change**

Grammaticalization is a gradual diachronic process of language change.<sup>19</sup> It does not occur abruptly, but goes through a series of small transitions that have a tendency to be cross-linguistically replicated. The order in which these transitions tend to occur is called a ‘chain’ or ‘pathway’ of change (e.g. Heine et al. 1991; Fischer et al. 2000). It should be noted, however, that grammaticalization is by no means deterministic. Changes do not have to occur and if a change takes place it does not have to go through all transition stages (Hopper & Traugott 2003:130-1). The term ‘cline’ is a theoretical abstraction, i.e. a generalization over actually

---

<sup>17</sup> [http://www.geisteswissenschaften.fu-berlin.de/en/v/Refining\\_Grammaticalization/](http://www.geisteswissenschaften.fu-berlin.de/en/v/Refining_Grammaticalization/) (17-04-2012).

<sup>18</sup> Moreover, the ambiguous notion of grammaticalization is often left undefined in the literature.

<sup>19</sup> Note that not all language change is grammaticalization. As here defined, grammaticalization (cf. subsection 2.1.1.5) is a composite change. This excludes borrowing, coinage or single changes in phonology/phonetics, morphology, semantics or discourse/pragmatics.

attested, similar grammaticalization chains. Clines are applicable to both synchronic and diachronic studies of grammaticalization processes, as quoted in (31).

- (31) The term “cline” is a metaphor for the empirical observation that cross-linguistically forms tend to undergo the same kinds of changes or have similar sets of relationships, in similar orders. [...] From a historical perspective, a cline is conceptualized as a natural “pathway” along which forms evolve, a schema which models the development of forms (see Andersen 2001). Synchronically, a cline can be thought of as a “continuum”: an arrangement of forms along an imaginary line at one end of which is a fuller form of some kind, perhaps “lexical,” and at the opposite end a compacted and reduced form, perhaps “grammatical” (Hopper & Traugott 2003:6).

A strong but controversial principle within grammaticalization studies is the so-called ‘unidirectionality hypothesis.’ This hypothesis postulates that grammaticalization processes are unidirectional in nature (Norde 2009:48-105). This means that a change can only take place in one direction, namely a (gradual) shift from lexical to grammatical status (and from grammatical to even more grammatical status). Hopper & Traugott (2003:16) define unidirectionality as follows:

- (32) The principle that has come to be known as unidirectionality is an assertion about the change

less grammatical > more grammatical

that is fundamental to grammaticalization. Unidirectionality is a strong hypothesis that is based on observations about change, observations that lead to the conclusion that grammatical forms do not in general move “uphill” to become lexical, whereas the reverse change, whereby grammatical forms are seen to have their origins in lexical forms, is widespread and well documented.

Unidirectionality applies to all linguistic levels: “semantics (fully referential > bleached/grammatical meaning; less subjective > more subjective), syntax (lexical > grammatical; less bound > more bound) and phonology (full phonological form > reduced phonological form)” (Fischer & Rosenbach 2000:20). The unidirectionality hypothesis is a contentious issue because of the claim that there are no exceptions to this principle, which implies that reverse changes from grammatical to lexical status (=degrammaticalization) do not exist.<sup>20</sup>

---

<sup>20</sup> A true reversal of a grammaticalization process would be ‘anti-grammaticalization’ (Haspelmath 2004), i.e. a change that leads from the endpoint to the starting point of a grammaticalization chain, and also shows the same intermediate stages.

These days it is generally accepted that processes of degrammaticalization, i.e., “a composite change whereby a gram in a specific context gains in autonomy or substance on more than one linguistic level (semantics, morphology, syntax, or phonology),” (Norde 2009:120) do exist, although cases of degrammaticalization are relatively rare compared to the many attested instances of grammaticalization (e.g. Heine & Kuteva 2002).

Traugott (2001:10) lists “some probably legitimate counterexamples” among which the development of infinitive *to*, Pennsylvania German *wotte* and cases of exaptation or regrammaticalization.<sup>21</sup> Campbell (2001:127-8) enumerates “some of the better known counterexamples.” These include, amongst others, the Irish first person plural suffix *-mid/-muid* which “was freed from being restricted as a bound suffix to become an independent pronoun, replacing the earlier independent first person plural pronoun *sinn* (Bybee et al. 1994:19-20),” Dutch *tig* (< *-tig*) ‘many times’ which was “segmented from forms such as *zestig* ‘60’, *negentig* ‘90’, etc. like English *umpteen*” (Norde 2006b) and the ‘s-genitive’ which developed “from a word-marking affix to a phrase-final clitic, with an intermediary stage at which the genitive was an inflectional phrase marker attached to smaller constituents” (Norde 2006a:201). Hence, the unidirectionality principle is best conceived of as a strong and robust tendency, rather than an absolute principle.

### 2.1.1.2 Mechanisms of grammaticalization

Grammaticalization is generally characterized by both formal and semantic changes. When it comes to formal change, the two general mechanisms by which grammaticalization takes place are REANALYSIS and ANALOGY (Hopper & Traugott 2003:63-4). These mechanisms are not specific or restricted to grammaticalization and are not subject to unidirectionality.

REANALYSIS is a covert operation that results in a new structural representation for a given linguistic string that is not immediately noticeable at its surface manifestation. There are different types of reanalysis that apply to different linguistic levels (hierarchical structure, constituent structure and category label), but there is no generally accepted terminology to refer to its subtypes (Harris & Campbell 1995; Newmeyer 1998; Rosenkvist 2004).

I will distinguish between three types of formal reanalysis, i) hierarchical reanalysis, ii) categorical reanalysis, and iii) constituent internal reanalysis, whose labels are based on Rosenkvist (2004:32-6). The first type is hierarchical reanalysis, which is a formal change that leads to a new structure with changed scope relations. Linguistic expressions that undergo this operation are so-called ‘extra-sentential elements’, such as various kinds of discourse markers, adjuncts or parentheses. An example is the development of adverbials with discourse marker function (Traugott & Dasher 2002: 152-189), like *in fact*. In (33a) *in fact* is a sentential adverb, but in “the rather artificial” example in (33b) *in fact* functions a discourse marker. The examples are taken from Traugott & Dasher (2002:157-8).

---

<sup>21</sup> The terms ‘exaptation’ (Lass 1990) and ‘regrammaticalization’ (Greenberg 1991) involve the recruitment and recycling of old material into new grammatical functions.

- (33) a. *Humanity is **in fact** absent.*  
 b. *Humanity is not often present. **In fact**, it/humanity is usually absent.*

A more complex example is the development of the discourse marker *anyway*, which Tabor & Traugott (1998) discuss in relation to structural scope expansion and grammaticalization. They see the rise of this discourse marker as the result of two successive scope increasing operations which they define in terms of c-command, i.e. a syntactic relation between nodes in a tree in generative grammar (Adger 2003:117). The development of *anyway* is repeated in (34),<sup>22</sup> in which (34a) represents the manner adverb *anyway* 'in any manner, by any means', (34b) exemplifies concessive *anyway* 'nonetheless' which appears to be restricted to the right periphery of a VP, and (34c) illustrates 'topic-resuming' *anyway* which is used to signal a return to a previous topic and serves as a discourse markers (Tabor & Traugott 1998: 255-257). Topic-resuming *anyway* can occur in many syntactic positions, but it must be surrounded by pauses wherever it occurs (ibid).

- (34) a. [IP Sche [VP excusy<sup>th</sup> [NP hir] [PP in any wey] ] ]  
 b. [IP She [VP [VP excuses [NP herself] ] anyway] ]  
 c. [E [IP She [VP excuses [NP herself] ] ] anyway]

They observe that “the transition from M[anner]-*anyway* to C[oncessive]-*anyway* involves scope increase, as does the shift from C[oncessive]-*anyway* to T[opic]R[esuming]-*anyway*.” That this is the case can be seen from the decrease in brackets around *anyway*.

The second type is categorical reanalysis, as in (35), which results in a change of category label. For example, the shift from a lexical category (e.g. main verb) to a grammatical category (e.g. (modal) auxiliary) is a categorical reanalysis. But also conversion or zero derivation is an instance of category shift (e.g. *(to) run* (verb) > *(a) run* (noun)).

- (35) [V] > [AUX]  
 OE *magan* 'to be strong/able' > PDE *may*

The third type, constituent internal reanalysis (Heine & Reh 1984) or segmental reanalysis (Heine et al 1991: Hopper & Traugott 1993: 41; Rosenkvist 2004:35), relates to a change in constituent boundaries. In example (36a), *be going to* is reduced to the fused form *be gonna* (Hopper & Traugott 2003:93). In (36b), a construction consisting of a head noun and a dependent noun is reanalyzed as a (complex) preposition and a head noun (ibid:51). Example

<sup>22</sup> The ‘E-node’ is a position for ‘expressive adjuncts.’

(36c) shows the univerbation of the verb phrase *may be* into adverbial *maybe*, and the Old English noun phrase *hlaf* ‘loaf’+ *weard* ‘guardian’ into the simple lexeme *lord*. The formalizations are taken from Rosenkvist (2004:32-6).

- |      |    |                                       |   |  |
|------|----|---------------------------------------|---|--|
| (36) | a. | [[A B] C]]                            | > | [[A [B C]]                               |
|      |    | [[ <i>be going</i> ] <i>to</i> ]]     | > | [[ <i>be</i> [ <i>gonna</i> ]]           |
|      | b. | [[A] [B C]]                           | > | [[A B] [C]]                              |
|      |    | [[ <i>back</i> ] <i>of the barn</i> ] | > | [[ <i>back of</i> ] [ <i>the barn</i> ]] |
|      | c. | [[A] [B]]                             | > | [AB]                                     |
|      |    | [[ <i>may</i> ] [ <i>be</i> ]]        | > | [ <i>maybe</i> ]                         |
|      |    | [[ <i>hlaf</i> ] [ <i>weard</i> ]]    | > | [ <i>lord</i> ]                          |

The different types of reanalysis may coincide. For example, constituent internal reanalysis may (e.g. *maybe*), but need not (e.g. *lord*), result in a change of category label. It may also result in altered scope relations, as in (36b). Similarly, a change in category label may be accompanied by hierarchical reanalysis, as is the case for adverbials that develop into discourse markers, as in (33) and (34).

ANALOGY, on the other hand, is an overt operation that refers to the attraction of extant forms to already existing constructions or paradigms. Examples are ‘rule-generalizations’ as is the case for Dutch *rozelaar* ‘rose bush’ which is modeled by analogy with *hazelaar* ‘hazel’ (VD). A paradigmatic example of analogy comes from the domain of plural formation. Given *stan-stanes* ‘stone-stones’, the form pair *shoe-shoen* was analogized to the form now used in present-day English, *shoe-shoes* (Hopper & Traugott 2003:64).

Reanalysis and analogy are essentially different mechanisms, yet they are also closely linked, for although only reanalysis can create new grammatical structures, they may become visible through analogy. The following citation, in (37), from Hopper & Traugott (2003:64) summarizes the differences and similarities between reanalysis and analogy.

- (37) In essence reanalysis and analogy involve innovation along different axes. Reanalysis operates along the “syntagmatic” axis of linear constituent structure. Analogy, by contrast, operates along the “paradigmatic” axis of options at any one constituent node (Jakobson and Halle 1956).

With respect to semantic change, the two basic mechanisms are METAPHOR and METONYMY. These are also not restricted to grammaticalization and independent of unidirectionality. From a structural perspective, they are related to the formal mechanisms analogy and reanalysis. Metaphorical change is similar to analogy in that it is a type of paradigmatic change, i.e. a change “whereby a word-sign used for a particular object or concept comes to be used for



another concept because of some element that these two concepts have in common” (Fischer & Rosenbach 2000:15).

METAPHOR in semantic change concerns concrete notions that are employed to express abstract concepts. Examples are body parts that come to denote spatial terms and configurations (*back* ‘body part’ > ‘behind’, *head* ‘body part’ > ahead ‘in front’), or mental activities that are conceptualized in terms of physical activities, such as English *grasp*, *comprehend*, German *fassen*, Dutch *begrijpen*, *(be)vatten*, *snappen*, which are all figurative abstractions of the literal sense of ‘grabbing (at) something’ (cf. Fischer & Rosenbach 2000; Hopper & Traugott 2003).

METONYMY is related to reanalysis in that it operates at the syntagmatic level. Metonymic association is not caused by similarity (as for metaphor) but by contiguity, that is, a sign is used that is indexically related to the substituted one. Examples of metonymy are expressions like *the press* referring to journalists or *a Munch* denoting a painting by the Norwegian artist Edvard Munch. In the modal domains of necessity and possibility there are many closely related and overlapping meanings. An example is the modal *must*, as in *He must be in the office*, that can be ambiguous between a deontic (i.e. He is obliged to be in the office) and epistemic reading (i.e. It must be the case that he is in the office), see Chapter 3. Here the inference is that when something is obligatory it has a high probability of occurring because it is mandatory.

The relation between metonymy and reanalysis and metaphor and analogy is summarized by Hopper & Traugott (2003:93), as cited in (38) below.

- (38) [M]etonymic and metaphorical inferencing are complementary, not mutually exclusive, processes at the pragmatic level that result from the dual mechanisms of reanalysis linked with the cognitive process of metaphor. Being a widespread process, broad cross-domain metaphorical analogizing is one of the contexts within which grammaticalization operates, but many actual instances of grammaticalization show that conventionalizing of the conceptual metonymies that arise in the syntagmatic flow of speech is the prime motivation for reanalysis in the early stages.

In short, the difference between metaphor and metonymy is that with metonymy a term from the same field (or context) is chosen, whereas with metaphor a term is substituted for a similar cognitive element from a different field or paradigm (Fischer & Rosenbach 2000:17).

### 2.1.1.3 Some examples of grammaticalization

A well-known and much-cited example of grammaticalization is the development of the phrase ‘*be going to*’ from purely locative marker into a marker of futurity (Fischer & Rosenbach 2000:3), as illustrated in (39) below.

- (39) a. *I am going (to Haarlem) to visit my aunt.*  
b. *I am going to marry (tomorrow).*

- c. *I am going to like it.*
- d. *It is going to rain.*
- e. *I am going to go there for sure.*
- f. *I'm gonna go.*

At first *be going to* has purely locative (and purposive) meaning, i.e. going to a place (Haarlem) for a certain purpose (to visit my aunt) in (39a). This meaning is extended to temporal meaning / futurity, i.e. a movement in time (tomorrow) in (39b) and intention / aim, i.e. purposive meaning in (39c). Little by little the structure becomes generally available to all kinds of expressions, i.e. general futurity in (39d) and (39e). The final stage, in (39f) is a colloquial and semi-auxiliary form *gonna* which is affected by phonetic reduction. Note that all stages co-exist up to now.

Another example of grammaticalization is the development of the preposition *ná* out of the verb 'give' *ná* in Ewe, an African language belonging to the Kwa branch of the Niger-Congo family, which is spoken in eastern Ghana, southern Togo, and southern Benin. In example (40a), *ná* is a verb meaning 'give', in (40b) it is ambiguous and may be interpreted alternatively as a verb ('give') or as a benefactive preposition ('for'), and in (40c) *ná* can only be interpreted as a preposition 'for', (Heine et al. (1991:1).

- (40) a.    me-**ná**            ga            kofi  
           1<sub>SG</sub>-give        money       Kofi  
           'I gave Kofi money'
- b.    me-ple            ßotrí **ná**        kofi  
           1<sub>SG</sub>-buy        door    give    Kofi  
           (a) 'I bought a door and give it to Kofi'  
           (b) 'I bought a door for Kofi'
- c.    me-wɔ dɔ'                    vévié **ná**        dodókɔ'        lá  
           1<sub>SG</sub>-do work            hard    give    exam            DEF  
           'I worked hard for the exam'

Other prototypical instances of grammaticalization include, inter alia, the development of the preposition *back* from a noun denoting a body part into a preposition meaning 'behind', the development of verbs like *may* or *must* from a lexical verb into a modal auxiliary, and the transformation of a general verb *bé* meaning 'to say' into a complementizer *bé* in Ewe. For more examples of grammaticalization see Heine & Kuteva (2002).

### 2.1.1.4 Properties of grammaticalization

In the literature, a vast number of prototypical properties of grammaticalization have been identified. However, the majority of these “parameters” (Lehmann 1995), “principles” (Hopper 1991) and “characteristics” (Brinton & Traugott 2005) turn out to be mere side effects rather than typical features of the change itself. Like the concept of grammaticalization, these ‘characteristic features’ are also prone to more than one interpretation

Lehmann’s (1995 [1982]:123) parameters of grammaticalization apply to ‘the autonomy of the linguistic sign’ from a paradigmatic and syntagmatic perspective, viz. the selection and combination of linguistic signs. Three facets of the linguistic sign are taken into account, namely its weight, cohesion, and variability. The values of these aspects may signal different degrees of grammaticalization (=loss of autonomy) of a given linguistic sign. The parameters are given in (41) below.

(41)		<b>paradigmatic</b>	<b>syntagmatic</b>
	<b>weight</b>	integrity	structural scope
	<b>cohesion</b>	paradigmaticity	bondedness
	<b>variability</b>	paradigmatic variability	syntagmatic variability

These six parameters are all gradient and linked to gradual primitive change (ibid:123, 164), as listed and described in (42).

- (42) a. **Integrity:** The weight of a sign, viewed paradigmatically, is its integrity, its substantial size, both on the semantic and phonological sides. The integrity of a linguistic sign may be affected by *attrition* (=loss of semantic and phonological substance, as well as morphosyntactic properties).
- b. **Structural scope:** The weight of a sign, viewed syntagmatically is its structural scope, that is the extent of the construction which it enters or helps to form. The structural scope of a linguistic sign may be subject to *condensation*, i.e. structural scope reduction. For example, an item that relates to a constituent of arbitrary complexity changes into an item that modifies a word or stem.
- c. **Paradigmaticity:** The cohesion of a sign with other signs in a paradigm [is] its paradigmaticity, that is, the degree to which it enters a paradigm, is integrated into it and dependent on it. The paradigmaticity of a linguistic sign may be regulated through *paradigmaticization*. For example, before grammaticalization takes place, an item may participate loosely in a semantic field, through the course of grammaticalization the item becomes part of a small, tightly integrated paradigm. Paradigmaticization includes shifts from major to minor category as well as integration into inflectional paradigm.

- d. **Bondedness:** The cohesion of a sign with other signs in a syntagm [is] its bondedness; this is the degree to which it depends on, or attaches to, such other signs. *Coalescence*, i.e. boundary loss and univerbation, leads to increased bondedness of a linguistic sign.
- e. **Paradigmatic variability:** The paradigmatic variability of a sign is the possibility of using other signs in its stead or of omitting it altogether (=freedom of choice). *Obligatorification* systematically constrains the paradigmatic variability of a sign whenever it becomes [grammatically] obligatory.
- f. **Syntagmatic variability:** The syntagmatic variability of a linguistic sign is the possibility of shifting it around in its construction. *Fixation* restricts the syntagmatic variability of a linguistic sign because the item occupies a fixed slot.

Lehmann's parameters have been widely applied in grammaticalization analyses, though not in a consistent manner, i.e. they are hardly ever all applied. Instead, only a subset of parameters is picked out that match a particular change. In the literature there is remarkably little discussion of the parameters themselves, and only a few linguists have criticized this taxonomic system (Detges & Waltereit 2002, Von Mengden 2008. See Norde (forthc.) for a discussion and revision of Lehmann's parameters in relation to grammaticalization and degrammaticalization.

The parameter of 'structural scope' has been discussed in the literature because both scope reduction and scope expansion are attested in grammaticalization. This suggests that scope is not a unidirectional parameter in grammaticalization. This relates to the two main approaches to grammaticalization (Traugott 2010a), viz. grammaticalization as reduction (and increased dependency) and grammaticalization as expansion. The parameter of *structural scope* and the correlated change of *condensation* apply to grammaticalization as reduction. Structural scope reduction is also in conflict with generative approaches to grammaticalization as elements that are merged higher up in the syntactic structure gain in syntactic scope.

Hopper (1991:22) describes five basic principles that can be used as heuristic devices to identify potential cases of grammaticalization. He stresses that these principles apply to language change in general, they are not restricted to grammaticalization. Even though the principles are not unique to grammaticalization, he wishes that "these principles will supplement those suggested by Lehmann in being characteristic of grammaticization not only at the later, more easily identifiable stages, but also at the incipient stages where *variable* phenomena occur, and where the question more cogently arises as to whether we might speak of grammaticization (ibid:21). His five principles of 'grammaticization' are given in (43) below.

- (43) a. **Layering:** Within a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers.<sup>23</sup>
- b. **Divergence:** When a lexical form undergoes grammaticization to clitic or affix, the original lexical form may remain as an autonomous element and undergo the same changes as ordinary lexical items.
- c. **Specialization:** Within a functional domain, at one stage a variety of forms with different semantic nuances may be possible; as grammaticization takes place, this variety of formal choices narrows and the smaller number of forms selected assume more general grammatical meanings.
- d. **Persistence:** When a form undergoes grammaticization from a lexical to a grammatical function, so long as it is grammatically viable some traces of its original lexical meanings tend to adhere to it, and detail of its lexical history may be reflected in constraints on its grammatical distribution.
- e. **De-categorialization:** Forms undergoing grammaticization tend to lose or neutralize the morphological markers and syntactic privileges characteristic of the full categories Noun and Verb, and to assume attributes characteristics of secondary categories such as Adjective, Participle, Preposition, etc.

The principles layering, divergence, persistence and specialization are all side effects of de-categorialization. That is, they come along with every shift from major to minor category.

In order to distinguish grammaticalization from the related process of lexicalization (Section 1.1.2), Brinton & Traugott (2005:110) provide a list of characteristics that are typical of grammaticalization. The features and their description are described in (44) below.

- (44) a. **Gradualness** refers to the fact that most changes occur in very small structural steps, typically with innovative uses coexisting alongside older ones (ibid:26). This can be illustrated by the following formulation, which characterizes changes from one state to another over time:
- $$A > A \sim B (> B)$$

---

<sup>23</sup>As correctly observed by Van Bogaert (2010:footnote 12) “what Hopper originally meant by ‘layering’ is ‘the prominent fact that very often more than one technique is available in a language to serve similar or even identical functions’ (1991:23), the term is now commonly reinterpreted as the synchronic coexistence of more and less grammaticalized manifestations of a given form (see Brems 2007; Brems et al. 2008; Aijmer 2009).” In the present study, layering is also used in the sense of synchronic co-existence of more and less grammaticalized variants of a given form.

Eventually the newly developed, innovative use may become the only possible use of a given form.

- b. **Unidirectionality** is an empirical hypothesis according to which changes in the reverse direction from grammaticalization are infrequent or even impossible (see Norde 2009: 48ff. for discussion). Both fusion (point c below) and coalescence (point d below) are factors in unidirectionality. Grammaticalization leads to more abstract, functional meaning (ibid:105)
- c. **Fusion** involves freezing and fixing of collocations, e.g., *take a {walk, bath, bite, fall, look, nap}*, (ibid).
- d. **Coalescence** is the reduction of phonological segments subsequent to fusion. In the case of grammaticalization it may lead to reduced forms (e.g., *want to* > *wanna*) and to regular patterns such as paradigms (ibid).
- e. **Demotivation**, idiomatization, or loss of semantic compositionality. These are also factors in unidirectionality. [I]n the case of grammaticalization it leads to more general and abstract grammatical meaning.
- f. **Metaphorization/metonymization** are both general kinds of semantic change (see Section 2.1.1.2). Metaphorization applies to conceptualization of structures across domains and metonymization relates to contiguity or association of structures within domains.
- g. **Decategorialization** (see Hopper 1991:21 in (43) above) is a defining characteristic of grammaticalization since it [is] the mechanism by which lexical items become functional.
- h. **Bleaching** is weakening of meaning through generalization, most especially loss of contentful meaning, at least in late stages of grammaticalization. Items that can undergo grammaticalization tend to have quite general meanings (e.g., terms for ‘thing,’ ‘go,’ ‘come,’ ‘behind’). It should be noted, however, that bleaching alone is a misleading concept: while content meaning may be reduced or even lost in grammaticalization, pragmatic and indexical meaning is added (Hopper & Traugott 2003:94-98).
- i. **Subjectification** is the anchoring of meaning in the speaker’s assessment of the situation. Since grammaticalization involves shifts toward more abstract, less referential, markers, the prime function of which is to represent the speaker’s

perspective on the situation or to get others to do things, it is necessarily the case that subjectification is characteristic of grammaticalization. Subjectification is, however, not unique to grammaticalization (cf. the development of speech act verb meanings of lexical items like *promise*) (Traugott & Dasher 2002). See also Section 2.1.4.

- j. **Productivity.** Items that grammaticalize become more productive in the sense that the grammaticalizing element occurs with increasingly large numbers of categories, i.e., with increasing type frequency. The shift is from a less to more productive pattern, (see Lehmann's "paradigmaticization" (1995:164) and Himmelmann's "host-class expansion" (2004:33).
- k. **Frequency:** Items that grammaticalize are used "in more contexts and for a larger set of lexical items"; therefore grammaticalizing items always become more token frequent than their source (Himmelmann 2004:37).
- l. **Typological generality.** Grammaticalization patterns tend to be cross-linguistically replicated (Bybee et al. 1994; Heine & Kuteva 2002) and may affect whole semantic classes; e.g., body parts often become case markers, verbs of intentionality become modals, verbs of motion become future markers.

There is some overlap between the proposed parameters, principles and characteristics. The notions of 'gradualness' and 'unidirectionality' seem to some extent redundant in the sense that they are implied in most definitions of grammaticalization. The suffix *-ization* denotes that grammaticalization is a gradual diachronic process. Unidirectionality is inherent in the definition in that there is a change from lexical to grammatical, and from grammatical to even more grammatical status. Decategorialization, as defined by Hopper and Brinton & Traugott, is a combined process that consists of a restricted unidirectional type of categorical reanalysis, i.e. from major to minor category, and subsequent loss of morphosyntactic properties (=attrition in Lehmann's terms).

#### 2.1.1.5 Towards a definition of grammaticalization

Taking the definitions and characterizations in (20) to (30), and the identified parameters (Lehmann 1995), principles (Hopper 1991) and characteristics (Brinton & Traugott 2005), as a point of departure, I will propose the following characterization of grammaticalization in (45) and Table 6 in which formal reanalysis and semantic reinterpretation are equally important. I distinguish between two types of grammaticalization, viz. 'primary grammaticalization' (=from lexical to grammatical status), and 'secondary grammaticalization' (=from grammatical to

(more) grammatical status).<sup>24</sup> Primary and secondary grammaticalization are abbreviated as Gzn1 and Gzn2, respectively. In Table 6, the sign ‘+’ stands for a key-defining property of grammaticalization, ‘-’ denotes that a certain feature does not apply to grammaticalization, and ‘(+)’ represents characteristics that may, but need not, be involved in grammaticalization.

- (45) Grammaticalization is a composite type of language change whereby lexical or already grammaticalized items, in certain linguistic contexts, undergo both semantic reinterpretation and formal reanalysis, as in (i). It is accompanied by a subset of correlated primitive changes and side effects, as in (ii) and (iii). Grammaticalization leads to a grammatical item, i.e. a linguistic item belonging to a minor category, with relational meaning, secondary status, the prime function of which is to regulate grammatical structure and grammatical relations.

i.	<b><u>Mechanisms in grammaticalization</u></b>	<b>Gzn1</b>	<b>Gzn2</b>
	<i>reanalysis</i>		
	<b>-categorical reanalysis</b>		
-	formal reanalysis from major to minor category	+	-
-	formal reanalysis from minor to minor category	-	+
	<i>reinterpretation</i>		
	<b>- metaphorization and/or metonymization</b>		
-	semantic reinterpretation from referential to relational meaning	+	-
-	semantic reinterpretation of relational meanings	-	+
ii.	<b><u>Primitive changes in grammaticalization</u></b>	<b>Gzn1</b>	<b>Gzn2</b>
	<i>-phonology/phonetics</i>		
-	loss of phonological/phonetic substance (attrition)	(+)	(+)
	<i>-morphology</i>		
-	loss of morphological compositionality <sup>25</sup> (fusion + coalescence)	(+)	(+)
-	loss of morphosyntactic properties (attrition)	+	(+)
	<i>-syntax</i>		
-	loss of syntactic variability (fixation)	(+)	(+)
-	loss of syntactic autonomy (integration)	(+)	(+)
	<i>-semantics</i>		
-	loss of semantic substance (bleaching)	+	+
-	loss of semantic compositionality (demotivation)	(+)	(+)
	<i>-discourse/pragmatics</i>		
-	gain of speaker’s perspective (subjectification), see Section 2.1.4	(+)	(+)

<sup>24</sup> This includes continued or advanced grammaticalization, as well as shifts between minor categories ‘intra-categorical shifts’, i.e. recategorization within the same domain. See also Joseph (2005) on ‘lateral shifts.’

<sup>25</sup> Compositionality applies to compositional forms only, not to monomorphemic or polysyllabic items.



iii.	<u>Side effects of grammaticalization</u>	Gzn1	Gzn2
-	paradigmaticization (=increase in paradigmaticity)	+	+
-	obligatorification (=decrease of paradigmatic variability)	(+)	+
-	condensation (=structural scope reduction)	+	+
-	layering (synchronic variation of a given form), divergence (split), specialization, persistence	+	+
-	productivity (=context expansion)	+	+
-	frequency (=increased type and token frequency)	+	+
-	typological generality (=cross-linguistic patterns)	+	(+)

**Table 6.** Mechanisms, primitive changes and side effects of grammaticalization.

Reanalysis and reinterpretation are abrupt changes, which may seem to be in conflict with the gradual nature of grammaticalization,<sup>26</sup> but since reanalysis and reinterpretation do not immediately manifest themselves at the surface, it only becomes noticeable after a while through side effects. That is, reanalysis and reinterpretation is not perceptible as abrupt transformations because the spread of newly formed structures and interpretations through the linguistic system, and hence the speech community, is gradual.

It is generally assumed that pragmatic-semantic change precedes formal change. Both in formal and semantic change there is a stage of indeterminacy. Semantic change consists of gradual contiguous steps that result in so-called ‘bridging contexts’ (cf. Evans & Wilkins 2000:550; Heine 2002:84) in which a linguistic item is ambiguous between two interpretations, i.e. the older and newly developed meaning. Note that these ‘bridging contexts’ are only ambiguous for the listener/reader, not for the speaker/writer. In formal change, category status may be indeterminate as long as a linguistic expression has not yet acquired all of the characteristics of its new category (cf. example (40) of Ewe *ná* for an instance of structural ambiguity). Hence, there is gradience,<sup>27</sup> both within and between grammatical categories, as well as semantic domains.

In sum, grammaticalization is a composite change that consists of formal reanalysis and semantic reinterpretation accompanied by a subset of correlated primitive changes at different linguistic levels (phonology/phonetics morphology, syntax, semantics, and discourse/pragmatics) in such a way that they collectively form the change called grammaticalization. Grammaticalizing and grammaticalized items show signs of ongoing change or that a change has occurred, these are so-called side effects or concomitants of grammaticalization. These are diagnostics that can be used as useful heuristics to identify potential cases of grammaticalization.

<sup>26</sup> Haspelmath (1998) claims that grammaticalization does not need reanalysis. According to him, an abrupt change like reanalysis is incompatible with the gradualness of grammaticalization. Instead, he sees grammaticalization itself as a mechanism of change.

<sup>27</sup> Gradience is the synchronic correlate of gradualness. It denotes the continuum between linguistic categories. See Traugott & Trousdale (2010) for a volume devoted to the relation between gradience, gradualness and grammaticalization.

### 2.1.2 Lexicalization

In grammaticalization studies, lexicalization is generally conceived of as the cover term for the study of the origin and rise of lexical items and the gradual diachronic change leading to lexical expressions. The term is, however, also frequently used in other linguistic frameworks with an entirely different meaning. Synchronically, the notion of lexicalization may refer to “the coding of conceptual categories” and “relates to the extent to which there are links between conceptual representation and syntax, and how the nature of such links may be formalized” (Brinton & Traugott 2005:18-9). For example, the lexeme *kill* is the representation of a more abstract structure as CAUSE BECOME NOT ALIVE (McCawley 1968).

Diachronically, there are basically three main views on what constitutes lexicalization, as listed in (46). In (46a), a very broad view is presented which includes various kinds of word formation processes as well as the subtypes in (46b) and (46c). However, within grammaticalization studies, lexicalization is generally distinguished from routine processes of word formation, as illustrated by (46a), and is seen as either fusion (46b), or fusion and processes of separation (46b+c). All examples are taken from Brinton & Traugott (2005).

- (46) a. **adoption into the lexicon:** various processes of word formation, such as compounding (e.g. *wallpaper*), derivation (*membership*), conversion (*run* (verb), *run* (noun)), clipping (e.g. *flu* < *influenza*), ellipsis (e.g. *canary* < bird from the Canary Islands), blending (e.g. *blog* < (*we*)*b*+ *log*), back formation (e.g. *orientate* < *orientation*), acronyms (e.g. *NATO*), loan translations (e.g. German *Fernseher* ‘television’), coinage or root creation (e.g. *zap*) or metalinguistic citation (e.g. the letter ‘*k*’).
- b. **fusion:** various kinds of univerbations resulting in morphologically and semantically opaque linguistic items/constructions (e.g. *lord* < OE *hlaf* ‘loaf’ + *weard* ‘guardian’).
- c. **separation:** clipping of bound morphemes resulting in semi-independent words whose meaning is not completely derivable or predictable (e.g. *ism*, *ology*).

The view in (46a) brings up the concept of ‘the lexicon’ as storage for lexical items. The lexicon is primarily a theoretical concept (as distinct from a dictionary), as described in (47a,b), and is often opposed to grammar (as distinct from a grammar book), as in (47c,d). Conceptions of the lexicon are theory dependent (cf. Brinton & Traugott 2005:3-5 for a discussion of different ideas about the lexicon).

- (47) a. A finite list of stored forms and possibilities for combining them (Brinton & Traugott 2005:9).

- b. The complete set of meaningful units in a language; the words, etc., as in a dictionary, but without the definitions (OED).
- c. [T]he set of categories, patterns, and organizing principles evidenced by language, most essentially abstract patterns of semantics, syntax, morphology, and phonology that at least in theory permit infinite combinations (Brinton & Traugott 2005:4).
- d. [I]nflexional forms or other means of indicating the relations of words in the sentence, and with the rules for employing these in accordance with established usage (OED).

The relation between grammar and lexicon has been conceptualized as a rigid distinction (e.g. the lexicon is a list of idiosyncratic items that is inserted into syntactic structures), or as a continuum that includes parallels between lexical and grammatical organization (e.g. the lexicon is multi-structured and includes not only highly idiosyncratic, but also more regular elements).

In grammaticalization studies the notion of ‘lexicon’ is often replaced by ‘inventory.’ The advantage of the broader notion of ‘inventory’ is that “[t]ypes of ‘adoption into the inventory’ can be differentiated as lexicalization (both in the broadest (46a) and in the narrow sense (46b)+c) KB) or grammaticalization, depending on the function of the adopted item, and shifts in function within the inventory can be accounted for (Brinton & Traugott 2005:90).

### **2.1.2.1 Characterizations of lexicalization**

In this section, different definitions of lexicalization that have been put forward in the literature will be discussed. In a broad sense, lexicalization refers to the adoption of lexical items into the lexicon or inventory, as described in (48), (49) and (50).

- (48) [T]he adoption of a word into the lexicon of a language as a usual formation that is stored in the lexicon and can be recalled from there for use (Bussmann 1996).
- (49) [A] process by which new linguistic entities, be it simple or complex words or just new senses, become conventionalized on the level of the lexicon (Blank 2001:1603).

Lexical items can be viewed from a holistic or componential approach. That is, whether they are treated as ‘unanalyzable wholes’ (Bloomfield 1933; Chomsky 1965), accessed ‘holistically’, i.e. “grasp the whole without consideration of the parts” (Lehmann 2002:2), or whether the conception of lexical items is based on minimal components of meaning (e.g. *boy* represents +HUMAN, -ADULT, +MALE e.g. Fillmore 1970), cf. Brinton & Traugott (2005:9ff). Lehmann explicitly mentions this point in his description of lexicalization in (50).

- (50) Lexicalization involves a holistic access to a unit, a renunciation of its internal analysis. Lehmann (2002:15). [I]t reduces the inner structure of a unit, shifting it into the inventory (ibid:18) [...].

For example, when a complex construction, like  $[XY]_Z$ , undergoes lexicalization it affects Z as a whole. The internal relations of Z become irregular and get lost. Note that on this view lexicalization necessarily involves internally complex units, and as a consequence “the coalescence of two grammatical morphemes must be called lexicalization” (ibid: 15).

Other views on lexicalization are construed in terms of ‘falling outside the productive rules of grammar’, as in (51) below.

- (51) Whenever a linguistic form falls outside the productive rules of grammar it becomes lexicalized (Antilla 1989 [1972]:151).

Lexicalization has also been characterized as a “reverse process” to grammaticalization (Kuryłowicz 1975 [1965]:52) or as (a type of) degrammaticalization (e.g. Ramat 2001; van der Auwera 2002), as cited in (52) and (53) respectively.

- (52) Contrary to a widespread opinion there are significant cases of items leaving the domain of the grammar and entering that of the lexicon- and lexicalization is indeed a subset of degrammaticalization (Ramat 2001:1).
- (53) Lexicalization is the making of a lexical item out of something other than a lexical item (van der Auwera 2002:21).

A characterization in terms of ‘fusion’ is provided by Himmelmann (2004). He distinguishes between two types of lexicalization, namely univerbation and fossilization, as stated in (54).

- (54) a. univerbation (the emergence of new lexical entries from collocations)
- b. fossilization (morphologically complex forms become unanalyzable wholes)

Brinton & Traugott (2005:96) define lexicalization as a gradual diachronic process by which a linguistic item becomes ‘more lexical’, as quoted in (55) below.

- (55) Lexicalization is the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or the word formation pattern. Over time there may be further loss of internal constituency and the item may become more lexical.

In addition to this definition, Brinton & Traugott (2005:110) list a number of properties that lexicalization has in common with grammaticalization. These will be discussed in the next section.

### 2.1.2.2 Properties of lexicalization

Brinton & Traugott's overview of the parallels between lexicalization and grammaticalization provides a good starting point for characterizing lexicalization and analyzing the grammaticalization-lexicalization interface (Section 2.2.1). The parallels and their brief descriptions are listed in (56) below. Recall that the same properties have been mentioned in relation to grammaticalization in (44).

- (56) a. **Gradualness** refers to the fact that most changes occur in very small structural steps, typically with innovative uses coexisting alongside older ones (ibid:26). This can be illustrated by the following formulation, which characterizes changes from one state to another over time:
- A > A~B (> B)
- Eventually only the newly developed, innovative use may become the only possible use of a given form.
- b. **Unidirectionality** in lexicalization leads to more contentful [=referential KB] meaning.
- c. **Fusion** in lexicalization involves the freezing and fixing of collocations, e.g., *strike a balance*, *hit the mark*.
- d. **Coalescence** is the reduction of phonological segments subsequent to fusion. In the case of lexicalization, coalescence tends to lead to phonological reductions and to irregular formations such as the result of phonogenesis (*alderman*) and phonologization (*sit-set*), see also Section 2.1.2.3.
- e. **Demotivation**, idiomatization, or loss of semantic compositionality. These are also factors in unidirectionality. In the case of lexicalization loss of compositionality tends to lead to increase in semantic specificity, contentfulness, and idiosyncrasy.
- f. **Metaphorization/metonymization**: these are both general mechanisms of semantic change, cf. Section 2.1.1.2. Metaphorization applies to conceptualization of structures across domains and metonymization relates to contiguity or association of structures within the same domain.

There are, however, some difficulties with this scheme of parallels between lexicalization and grammaticalization. First, there is no property typical of lexicalization that is not typical of grammaticalization too. This reduces lexicalization to a process that consists of a particular subset of grammaticalization criteria so that it cannot be an independent process. Second, the “minimal parallels” are not parallels, but distinctive properties for lexicalization and grammaticalization.

### 2.1.2.3 Some examples of lexicalization

According to Brinton & Traugott (2005:98), the following processes and items in (57) qualify as instances of lexicalization, because “the output is new or modified forms which are semantically contentful/ “lexical,” not functional/indexical/“grammatical.” Semantically, the items differ from their immediate sources by being more idiomatic and less compositional; morphophonologically they are more fused; with respect to productivity, their ability to collocate with host-classes becomes reduced.”

- (57) a. **fused syntactic phrases**, accompanied by idiomatization (*bread-and-butter* ‘necessities of life’), and sometimes undergoing morphophonological change (*handicap* < *hand in cap*).
- b. **fused (opaque KB) compounds**, such as *mildew* < OE *mele* ‘honey’ + *deaw* ‘dew’.
- c. **phonogenesis**, i.e. “the process whereby new syntagmatic phonological segments are created out of old morphemes” (ibid:52), such as *handiwork* < OE *handgeweorc*, *mayor* < Lat. *major* ‘great’ + ‘or’ COMP.
- d. **phonologization**, i.e., “phonologically conditioned alternations may eventually split into new phonemes as a result of erosion of segments” (ibid: 54), such as *drink/drench*. It is a paradigmatic process affecting the inventory of phonemes in a language.
- e. **creation of semantic, non-category-changing affixes**, such as *-hood* < OE *had* ‘rank.’

In the present study, phonogenesis and phonologization are excluded from lexicalization (and grammaticalization) as these are (single) phonological changes, not composite changes.

### 2.1.2.4 Towards a definition of lexicalization

Taking the proposed definitions in (46) to (55) and the identified properties as a starting point, I propose the following characterization of lexicalization in (58) and Table 7, in which formal and

semantic change are equally important. Accompanying primitive changes may manifest themselves at several linguistic levels, i.e. phonology/phonetics, morphology, syntax, semantics and discourse/pragmatics (cf. Bauer 1983; Wischer 2000). In the current study, the definition of lexicalization includes the subtypes fusion (lexicalization I) and separation (lexicalization II). These are abbreviated as Lxn1 and Lxn2, respectively. Regular processes of word formation which are largely transparent, as in (46a) are excluded from lexicalization. In Table 7, the sign ‘+’ stands for a key-defining property of lexicalization, ‘-’ denotes that a certain feature does not apply to lexicalization, and ‘(+)’ represents characteristics that may, but need not, be involved in lexicalization.

- (58) Lexicalization is a composite type of language change whereby (part of) a complex lexeme or (part of) a syntagm, in certain linguistic contexts, undergoes both formal reanalysis and semantic reinterpretation, as in (i). It is accompanied by a subset of correlated primitive changes and side effects, as in (ii) and (iii). Lexicalization leads to a lexical item, i.e. a linguistic item belonging to a major category, with referential meaning, primary status, and which may convey the main point of linguistic message.

i.	<b><u>Mechanisms in lexicalization</u></b>	<b>Lxn1</b>	<b>Lxn2</b>
	<b><i>reanalysis</i></b>		
	<b>-constituent internal reanalysis</b>		
-	formal reanalysis from syntagm > lexeme (=fusion)	+	-
e.g.	<i>handicap</i> (< <i>hand in the cap</i> ) <i>either</i> (< OE <i>a-hwæper</i> ‘each of two’)		
-	formal reanalysis from complex lexeme > simple lexeme (=fusion)	+	-
e.g.	<i>mermaid</i> (< OE <i>mere</i> ‘sea’ + <i>mægd(en)</i> ‘maiden’) <i>cobweb</i> (< OE ( <i>atter</i> ) <i>coppe</i> ‘spider’+ <i>web</i> ‘web’)		
-	formal reanalysis from bound morpheme > semi-independent word (=separation)	-	+
e.g.	<i>ism</i> (< X-ism e.g. expressionism, surrealism, etc.) <i>ization</i> (X-ization e.g. lexicalization, grammaticalization, etc.)		
	<b><i>reinterpretation</i></b>		
	<b>-metaphorization and/or metonymization</b>		
-	semantic reinterpretation of referential meanings (=generalization or specialization)	+	-
-	semantic reinterpretation from referential/relational to referential meaning (=concretion)	-	+

ii.	<b><u>Primitive changes in lexicalization</u></b>	<b>Lxn1</b>	<b>Lxn2</b>
	<b>- phonology/phonetics</b>		
-	loss of phonological/phonetic substance (attrition)	(+)	-
	<b>-morphology</b>		
-	loss of morphosyntactic properties (attrition)	-	-
-	loss of morphological compositionality (fusion + coalescence)	+	-
	<b>-syntax</b>		
-	loss of syntactic variability (fixation)	-	-
-	loss of syntactic autonomy (integration)	-	-
	<b>-semantics</b>		
-	loss of semantic substance (bleaching)	-	-
-	loss of semantic compositionality (demotivation)	+	-
	<b>-discourse/pragmatics</b>		
-	gain of speaker's perspective (subjectification), see Section 2.1.4.	(+)	(+)
iii.	<b><u>Side effects of lexicalization</u></b>	<b>Lxn1</b>	<b>Lxn2</b>
-	paradigmaticization (=increase in paradigmaticity)	-	+
-	obligatorification (=decrease of paradigmatic variability)	-	-
-	condensation (=structural scope reduction)	-	-
-	layering (synchronic variation of a given form), divergence (split), specialization, persistence	+	+
-	productivity (=context expansion)	-	+
-	frequency (=increased type and token frequency)	-	+
-	typological generality (=cross-linguistic replicated patterns)	-	(+)

**Table 7.** Mechanisms, primitive changes and side effects of lexicalization.

Note that most of these properties are the opposite of changes involved in grammaticalization, see (45). This does not mean that lexicalization is the reverse of grammaticalization. A true reversal of lexicalization would be anti-lexicalization, a change whereby exactly the same steps are repeated in inverse direction. The opposite change to lexicalization, ‘delexicalization’, is folk etymology. This pertains to the modification of “a lexical item (which may be historically complex, but has become opaque to speakers) is reanalyzed and given a morphological structure that it did not have before and that appears to be at least partially more transparent” (Brinton & Traugott 2005:83). Two examples from English are Old English *samblind* ‘half blind’ which became *sandblind*, and Old English *weddlac* ‘pledge-gift’ which was reanalyzed as *wedlock* (ibid).

In sum, lexicalization includes two subtypes: fusion and separation. These are composite changes that consist of formal reanalysis and semantic reinterpretation and are accompanied by several correlated primitive changes in such a way that they collectively form the composite



change called lexicalization. The side effects of these changes may be used to identify instances of lexicalization I and II.

### 2.1.3 Pragmaticalization

Pragmaticalization can be thought of as the study of the origin and rise of discourse markers, as well as the gradual diachronic change leading to discourse markers. The term pragmaticalization was introduced by Erman & Kotsinas (1993) in a study on Swedish *ba'* and its close correspondent *you know* in English. Pragmaticalization is a much-debated concept, but there are remarkably few attempts to come up with an explicit definition. As observed by Dieward (2011:373), pragmaticalization has not yet been defined in its own right, but has been employed to make a distinction between the domains of grammar and discourse. Hence, there is no consensus about a definition, let alone the status of pragmaticalization.

It is generally acknowledged that there is a domain of discourse, as distinct from grammar and lexicon or inventory, which relates to larger linguistic units than lexical or grammatical expressions in their respective contexts, as described in (59) and (60) below.

- (59) A connected series of utterances by which meaning is communicated, esp. forming a unit for analysis; spoken or written communication regarded as consisting of such utterances (OED).
- (60) The organization of language above the sentence or above the clause, [...] larger linguistic units, such as conversational exchanges or written texts. [L]anguage use in social contexts, and in particular with interaction or dialogue between speakers (Stubbs: 1983a:1).

Discourse markers, such as *oh*, *well*, *and*, *but*, *or*, *so*, *because*, *now*, *then*, *I mean*, *y'know* (Schiffrin 1987:31), are those elements that regulate and organize discourse structure. They are primarily concerned with the communicative aspects of language. Discourse markers originate in either lexical or grammatical items, which cause some items to assume dual statuses (e.g. *and*, *but*). That is, they may function both as a grammatical marker and as a discourse marker.

The development of discourse markers is a contentious issue in grammaticalization studies as it generally conforms to some well-known grammaticalization criteria while at the same time violating other basic principles of grammaticalization. Especially the optionality and multipositionality of discourse markers pose serious problems for traditional grammaticalization analyses. Discourse markers are linguistic items that operate at the discourse level, they do not belong to the core grammar. That is, they mark discourse structure, not grammatical structure. Because of this, their development has been termed 'pragmaticalization.' The status of pragmaticalization is controversial because it is not a generally accepted type of language change. Most linguists are reluctant to assign a special status to the rise of discourse markers,

while at the same time admitting that the development of discourse markers is by no means a prototypical case of grammaticalization.

### 2.1.3.1 Views on pragmaticalization

Different proposals have been made in order to account for the development and status of discourse markers, and how to accommodate this within the framework of grammaticalization studies. According to Brinton (1996) and Brinton & Traugott (2005) the development of discourse markers falls within the realm of (a broad characterization of) grammaticalization, Diwald (2011) defines pragmaticalization as “grammaticalization of discourse functions.” They have also been termed a distinct process called ‘pragmaticalization’ (Aijmer 1997) because of their interpersonal meanings and uses. Some of these views are fundamentally different, for others it just seems to be a matter of terminology.

In sum, there are basically four positions with regard to the relationship between grammaticalization and discourse, as summarized by Ocampo (2006:316-7), repeated in (61) below.

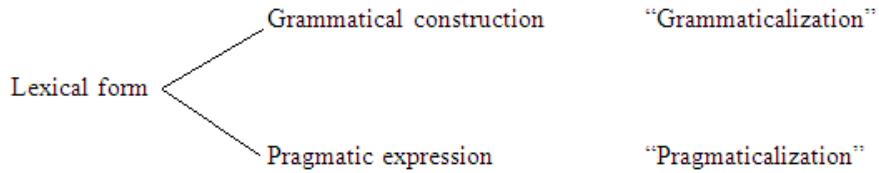
- (61) Some authors include movement towards discourse within the process of grammaticalization without any change in this notion (Onodera 1995, Brinton 1996, Pinto de Lima 2002). Another position is to expand the notions of grammar and grammaticalization so that they encompass movement towards discourse (Traugott 1995, Lenker 2000). The third position is to postulate two subtypes: *grammaticalization I* movement towards the morpheme, and *grammaticalization II*, movement towards discourse (Wischer 2000). A fourth position, which may be considered a variant of the previous one, is to postulate the notion of *pragmaticalization* (Aijmer 1997, Erman and Kotsinas 1993, Günthner and Mutz 2004).

The present study adheres to the fourth position. Pragmaticalization is considered to be genuinely different from both grammaticalization and lexicalization, and as such taken to be particular type of language change that involves the rise of discourse markers.

### 2.1.3.2 Properties of pragmaticalization

Pragmaticalization is well-known through the work of Aijmer (1997) on English *I think*. She makes a distinction between grammaticalization and pragmaticalization, which is visualized in Figure 1 below (ibid:2). According to Aijmer, the ‘truth-conditional criterion’ is of overriding importance for distinguishing between grammatical(ized) and pragmatic(alized) elements. That is, “elements which cannot be analysed in terms of truth are pragmatic or pragmaticalized.” According to her, other criteria are non-sufficient since they either pick out a sub-group of pragmatic(alized) expressions or characterize too large a class of adverbials (ibid:3). Note that on this view, all modal expressions are instances of pragmaticalization (cf. the description of modality in (3)). I do not find the ‘truth-conditional criterion’ a distinctive and sufficient

criterion for pragmaticalization, since modality may also be expressed grammatically.



**Figure 1.** Grammaticalization vs Pragmaticalization (Aijmer)

This characterization is primarily a semantic description, and as such it is hard to distinguish pragmaticalization from processes of (inter)subjectification (Section 2.1.4). Aijmer notes that “discourse markers such as *you know*, *you see* etc. are typically “pragmaticalized” since they involve the speaker’s attitude to the hearer.” It remains, however, obscure how pragmaticalization manifests itself formally.

More recently, the notion of pragmaticalization has been described more explicitly, for example by Günthner & Mutz (2004), Dostie (2009) and Beeching (2009), as shown in (62)-(64). However, like for lexicalization and grammaticalization there are many borderline cases that cannot straightforwardly be termed an instance pragmaticalization, I will return to this point in Section 2.2.2 and 2.2.3.

- (62) “Pragmaticalization”, then, would be the term to denominate the kind of diachronic change where elements [...] assume functions on the discourse-pragmatic level (cf. Erman & Kotsinas 1993; Aijmer 1997). [...] This type of change which leads to discourse and pragmatic markers, to elements which organize, structure, and contextualize discourse with respect to discourse-pragmatic concerns and not with respect to sentence-grammatical concerns (e.g. congruence, binding), contradicts classical grammaticalization [i.e. grammaticalization as reduction and increased dependency KB]. Whereas morphologization and syntacticization are classical instances of a grammaticalization process, the process of “pragmaticalization” as described and illustrated in this article, has to be regarded as a somewhat different (sub)type of linguistic (diachronic) change (Günthner & Mutz 2004:98-99).
- (63) The term [pragmaticalization] refers to a process of linguistic change in which a full lexical item (noun, verb, adjective or adverb) or grammatical item (coordinator, subordinator, etc.) changes category and status and becomes a pragmatic item, that is, an item which is not fully integrated into the syntactic structure of the utterance and which has a textual or interpersonal meaning (Dostie 2009:203).

- (64) Pragmaticalization refers to the process whereby a lexical/grammatical item develops uses which are conversational (related to discourse strategies) rather than propositional. [It] leads from M1 [meaning 1] to M2 [meaning 2], from a lexical or grammatical item to a semantically relatively bleached pragmatic particle, is generally considered to occur through the semanticization of “invited inferences” (Traugott and Dasher, 2002) and to be a gradual and unidirectional process. [...] [T]he nature of the pragmaticalization process which leads from M1 to M2 is a ticklish process, as the semantic change may be lengthy and remain for centuries at the M1/M2 stage characterized by polysemy and pragmatic ambiguity” (Beeching 2009:83).

Though pragmaticalization is a controversial concept in grammaticalization studies, I think there are good reasons to postulate a process of pragmaticalization as distinct from grammaticalization for the rise of linguistic items that operate at the discourse level (see Section 2.2.2 on the grammaticalization-pragmaticalization interface). Discourse markers conflict with standard definitions of grammaticalization and lexicalization. Discourse markers are not new contentful forms, nor do they belong to core grammar, but they are pragmatically required in order to organize discourse (cf. Diewald’s 2011:368 ‘communicative obligatoriness’).

### 2.1.3.3 Some examples of pragmaticalization

Since pragmaticalization is not a generally accepted type of language change, it is hard to come up with some prototypical, non-controversial examples. I will list some expressions that have been described as instances of pragmaticalization. Most of these examples concern various types of discourse markers in different languages, as in (65a), but Günthner & Mutz also consider modifying suffixes to be instances of pragmaticalization, as in (65b).

- (65) a. discourse markers, such as *you see*, *you know* in English (Aijmer 1997), *obwohl*, *wobei* in German (Günthner & Mutz 2004), and *coudon*, *écoute*, *pis*, *alors*, (*ça*) *fait que* in Quebec French (Dostie 2009).
- b. modifying suffixes in Italian, i.e. *-ino*, *-etto*, *-uccio*, *-one*, *-otto*, *-accio* (Günthner & Mutz 2004).<sup>28</sup>

Note that many expressions that have been classified as non-standard cases of either lexicalization or grammaticalization in traditional grammaticalization analyses, would count as instances of pragmaticalization in the present study (e.g. the development of ‘pragmatic markers’, e.g. *look*-forms, including *look’ee* (*looky*), *lookit*, (*now*) *look* (*here*) also *lookyhere*, *look-a-here*, which Brinton (2001) analyses as an instance of grammaticalization, though not a prototypical one).

<sup>28</sup> On my view, see Section 2.1.3.4 Towards a definition of pragmaticalization and 2.1.4 on (inter)subjectification, this is not a case of pragmaticalization but would be an instance of grammaticalization + subjectification.

### 2.1.3.4 Towards a definition of pragmaticalization

Pragmaticalization has not been as thoroughly studied as grammaticalization, lexicalization and their interrelations (cf. the previous sections). In order to sketch its profile, I will summarize and combine properties that are mentioned in the literature and make some first steps into creating a unified, principled account of pragmaticalization in such a way that it can be compared to lexicalization and grammaticalization. Taking the proposed definitions in (62)-(64) and the identified characteristics as a starting point, I propose the following definition of pragmaticalization in (66) and Table 8 in which formal and semantic change are equally important. In Table 8, the sign ‘+’ stands for a key-defining property of pragmaticalization, ‘-’ denotes that a certain feature does not apply to pragmaticalization, and ‘(+)’ represents characteristics that may, but need not, be involved in pragmaticalization.

- (66) Pragmaticalization is a composite type of language change, whereby lexical or grammatical expressions, in certain linguistic contexts, undergo both semantic reinterpretation and formal reanalysis, as in (i). It is accompanied by a subset of correlated primitive changes and side effects, as in (ii) and (iii). Pragmaticalization leads to a discourse marker, i.e. a linguistic item with conversational meaning, extra-propositional status, the prime function of which is to organize discourse structure.

<b>i.</b>	<b><u>Mechanisms in pragmaticalization</u></b>	
	<i>reanalysis</i>	
	<b>-hierarchical reanalysis</b>	
-	formal reanalysis from propositional > extra-propositional status <sup>29</sup>	+
e.g.	[I think (that) X] > [I think, [X]] / [[X], I think]	
	<i>reinterpretation</i>	
	<b>-metaphorization and/or metonymization</b>	
-	semantic reinterpretation from referential/relational meaning to conversational meaning (= (inter)subjectification), see Section 2.1.4.	+
<b>ii.</b>	<b><u>Primitive changes in pragmaticalization</u></b>	
	<b>-phonology/phonetics</b>	
-	loss of phonological/phonetic substance (attrition)	(+)
	<b>-morphology</b>	
-	loss of morphosyntactic properties (attrition)	(+)
-	loss of morphological compositionality (fusion + coalescence)	(+)

<sup>29</sup> i.e. parenthetical status of structural elements (e.g. insertions/afterthoughts) which have no grammatical connection to the clause to which they attach.

	<b>-syntax</b>	
-	loss of syntactic variability (fixation)	-
-	loss of syntactic autonomy (integration)	-
	<b>-semantics</b>	
-	loss of semantic substance (bleaching)	+
-	loss of semantic compositionality (demotivation)	(+)
	<b>-discourse/pragmatics</b>	
-	gain of speaker's perspective (subjectification), see Section 2.1.4.	+
<b>iii.</b>	<b><u>Side effects of pragmaticalization</u></b>	
-	paradigmaticization (=increase in paradigmaticity)	(+)
-	obligatorification (=decrease of paradigmatic variability)	-
-	condensation (=structural scope reduction)	-
-	layering (synchronic variation of a given form), divergence (split), specialization, persistence	+
-	productivity (=context expansion)	+
-	frequency (=increased type and token frequency)	+
-	typological generality (=cross-linguistic patterns)	(+)

**Table 8.** Mechanisms, primitive changes and side effects of pragmaticalization.

In sum, pragmaticalization is a composite change that consists of formal reanalysis and semantic reinterpretation, which is accompanied by a subset of correlated primitive changes at several linguistic levels. The side effects of these changes may be used to identify instances of pragmaticalization. Note that pragmaticalization is not the same as (inter)subjectification, or grammaticalization accompanied by (inter)subjectification. These differences will be discussed in the next section.

#### 2.1.4 Subjectification and intersubjectification

The terms subjectification and intersubjectification are often discussed in relation to grammaticalization and pragmaticalization. Subjectification is, however, not restricted to grammaticalization and pragmaticalization, as there are also cases of subjectification in lexicalization. It is important to keep in mind that (inter)subjectification is a specific, metonymic type of semantic change, not a composite change like lexicalization, grammaticalization and pragmaticalization. Subjectification and intersubjectification may accompany these different types of language change, but are not dependent on any of these types of language change.

Like the concepts of grammaticalization, lexicalization and pragmaticalization, (inter)subjectification is no exception to terminological fuzziness. The notions of (inter)subjectivity and their diachronic correlates (inter)subjectification are elusive concepts that have been conceptualized in various ways. The terms suffer from general vagueness concerning the nature of (inter)subjectivity, the heterogeneity of the phenomena grouped under the label of

(inter)subjectification and a lack of (formal) criteria to identify and measure instances of (inter)subjectification. Broadly speaking, subjectivity is the presence, involvement or reference to the speaker in language, whereas intersubjectivity concerns reference to speaker-hearer/writer-reader interaction.

In a general sense all language is (inter)subjective because it is ultimately produced by a speaker/writer in some communicative context. This observation has been made by several linguists (Bréal 1964 [1900]; Benveniste 1966; Lyons 1977, 1982) and is what de Smet & Verstraete (2006:370) call (general) ‘pragmatic subjectivity’, as described in (67).

- (67) [A]ll language use passes through a speaker and is, as such, speaker-related. Every choice of words figures in the speaker’s larger discourse and is indicative of the speaker’s attitudes and rhetorical strategies. In this sense, subjectivity relates to the use of an expression, not to the conceptual content carried by that expression.

What is of interest in the present study is the semantic process of (inter)subjectification, its nature, manifestations and subtypes. Hence, ‘general pragmatic subjectivity’ as in (67) should be kept apart from the specific metonymic type of semantic process itself.

Many different views on subjectification and intersubjectification have been voiced in the literature. There are three main traditions when it comes to studies of (inter)subjectification, namely the views of Langacker, Traugott and Nuyts. I will briefly discuss these different approaches in 2.1.4.1-3 below.

#### 2.1.4.1 Langacker

Langacker uses the terms subjectivity and subjectification in a rather technical sense, which is not always easily transferable to linguistic analysis. The idea is that an utterance is maximally subjective when there is no overt reference to the speaker. This means that the speaker is implicitly present or offstage. A maximally objective utterance contains overt reference to the speaker who is explicitly present or onstage.

An example of an objective linguistic item is the prepositional phrase *next to me*, as in (68a), because the speaker is explicitly mentioned in the utterance, i.e. *me*. The demonstrative *this*, as in (68b), is an instance of a subjective linguistic item because “proximity to the speaker is part of the meaning of the demonstrative *this*, but the speaker is not mentioned at all” (De Smet & Verstraete 2006:369).

In both examples the interpretation of the utterance needs speaker reference, but the coding of speaker involvement is formalized in a different way. The examples in (68) are adapted from Langacker (1985:118-9, 1990:12-13) by De Smet & Verstraete (2006:369).

- (68) a.     *The man **next to me** is James.*  
      b.     ***This** man is James.*

To illustrate the concept of subjectivity as Langacker defines it, compare the examples in (69) below (Langacker 1985:115):

- (69) a. *Jill is sitting **across** the table **from** Robyn.*  
b. *Jill is sitting **across** the table.*

In example (69a), the configuration does not need to include the speaker as it states how the persons are arranged with respect to each other. Example (69b), however, needs to include the speaker because “in the absence of a *from*-phrase, *across* is interpreted as referring to the spatial relation between the subject of the clause and the speaker” (De Smet & Verstraete 2006:366). Because of this, (69b) can be considered the subjective counterpart of (69a). Subjectification, then, is “the shift from a relatively objective construal of some entity to a more subjective one” (Langacker 1999:299). See Athanasiadou et al. (2006) for a collection of studies into subjectivity and subjectification undertaken from a Langackerian perspective.

#### 2.1.4.2 Traugott

For Traugott (inter)subjectivity and (inter)subjectification refer to semantic processes whereby a linguistic element develops new senses that involve speaker-reference or speaker-perspective. Expressions of subjectivity “index speaker attitude or viewpoint” and expressions of intersubjectivity mark “speaker’s attention to addressee self-image” (Traugott 2010:32).

The semasiological processes by which these meanings arise are subjectification and intersubjectification. Subjectification relates to meanings that “come to express grounding in the speaker/writer’s perspective explicitly”. Intersubjectification concerns meanings that “come to express grounding in the relationship between speaker/writer and addressee/reader explicitly” (Traugott & Dasher 2002:6). In Traugott’s view, the diachronic relatedness of the two concepts is central, that is, intersubjectification follows and originates in subjectification.

In essence, (inter)subjectification is semanticization of pragmatic implicatures/values. It requires the new (inter)subjective meanings to be conventionally coded by the forms, with new form-meaning pairs as a result. Hence, it is important to keep in mind that “pragmatic strengthening of subjective meanings is without question a pre-condition for subjectification, subjectification itself is not pragmaticalization,<sup>30</sup> but semanticization” (Traugott 2010:35).

Traugott’s approach differs from Langacker in the sense that Traugott defines subjectivity in terms of degrees of speaker-relatedness, whereas Langacker’s definition is based on implicit/explicit speaker-reference. See Davidse et al. (2010) for a compilation of studies into (inter)subjectivity and (inter)subjectification undertaken from a Traugottian point of view.

---

<sup>30</sup> Note that the use of the term ‘pragmaticalization’ differs from the one used in the present study; here it refers to “strengthening of pragmatic meanings.”



### 2.1.4.3 Nuyts

Nuyts' (e.g. 2001a) view on subjectivity has its origin in the (alleged) distinction between objective and subjective epistemic modality – a division that is often encountered in the literature on epistemic modality. According to him, subjectivity is an evidential dimension in epistemic modal expressions which he defines in terms of personal (subjectivity) versus shared responsibility (intersubjectivity), as stated in the citation in (70) below, see Section 1.1 on (epistemic) modality and evidentiality.

- (70) The dimension of subjectivity, as it is generally assumed to be present in epistemic modal expressions, is actually (in principle) an independent evidential-like qualificational category. This dimension of subjectivity is expressed in some, but not in other epistemic expression types. Further, it uses linguistic resources more or less separate from (though not independent of) the epistemic forms. The dimension of subjectivity should not be defined in the traditional terms of the quality of the evidence for an (epistemically qualified) state of affairs, but rather in terms of the question whether the evidence (and the conclusion drawn from it) is only available to the speaker or is rather more widely known (including to the hearer), (Nuyts 2001a:398-9).

This idea of subjectivity contrasts with both Langacker and Traugott's notion of subjectivity and subjectification. As observed by Nuyts, applying his own and Langacker's notion of subjectivity to the same type of linguistic items yields different outcomes. For example, Langacker considers the English modals to be highly subjectified as they are 'grounding predications', but in Nuyts' use of the notion of subjectivity, the modal auxiliaries are not subjective at all because they only receive "a (non-)subjective interpretation when the context imposes this – that is, when the meaning of the modal auxiliary unavoidably 'absorbs' the flavor of those contexts" (ibid:392).

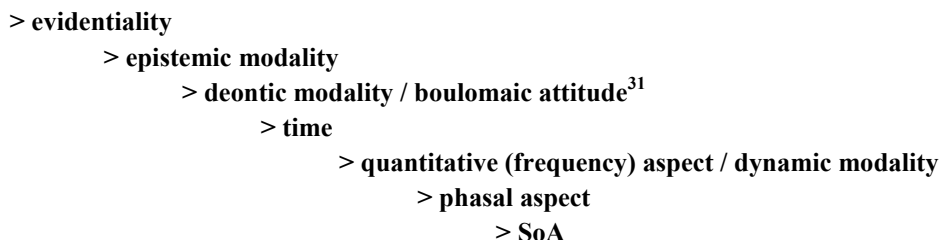
The mental state predicates are highly subjective in his analysis because they "systematically express subjectivity. Thus, they typically and predominantly occur in contexts in which the speaker voices personal opinions, very often about topics in the realm of strictly individual experiences or concerns, or also in contexts involving antagonism between the views of speaker and hearer" (ibid:390-1). However, mental state predicates "are clearly much less grounded, hence much less subjective than the modal auxiliaries in Langacker's concept" (ibid:392).

Though subjectivity in the sense of Nuyts also differs from Traugott's understanding of subjectivity, these two views can be related to each other. Nuyts (ibid) describes the differences and similarities between these approaches as follows in (71).

- (71) The present concept of the dimension of subjectivity [...] is much narrower than Traugott's notion of subjectification, but they are related to the extent that the present dimension is situated very high in the hierarchy in a layered model of the clause. That is,

the current notion of subjectivity is very high on the scale of subjectivity in Traugott's sense.

The diachronic correlate of subjectivity, subjectification, is represented as "climbing up in the hierarchy of qualificational categories" (Nuyts 2012), as exemplified in Figure 2 below.



**Figure 2.** The hierarchy of qualificational categories (Nuyts).

The concept of intersubjectification, then, pertains to the phenomenon that a linguistic item/expression is "leaving the hierarchy in Figure 2 to assume an interaction management function, e.g. as illocutionary marker, hedging device, sentence connector" (ibid).

#### **2.1.4.4 (Inter)subjectivity and (inter)subjectification in the present study**

The three main views on (inter) subjectification differ considerably from each other. Langacker views subjectivity in light of the construal of an event, i.e. whether speaker reference is explicitly present (onstage) or not (offstage). Traugott sees (inter)subjectification as a process of semantic reanalysis through which expressions of speaker-perspective or speaker-addressee interaction arise, and Nuyts characterizes subjectification as climbing up in the hierarchy of qualifications and intersubjectification as leaving the qualificational hierarchy. In this study, the general perspective on (inter)subjectivity and (inter)subjectification is in line with (inter)subjectification in a Traugottian sense (e.g. 2010).

The basic idea behind subjectivity and subjectification goes back to the work of Halliday & Hassan (1976), in which a distinction is made between three different linguistic components, as shown in (72) below.

(72)    ideational                      textual                      interpersonal

Traugott (1982) observed a diachronic relation between these layers, which she renamed 'propositional', 'textual', and 'expressive', as illustrated by the cline in (73).

<sup>31</sup> Boulomaic attitude concerns the degree of 'likability' of the state of affairs (cf. Dutch *Ik mag graag fietsen* 'I like to bike')

(73) propositional > ((textual) > (expressive))

The notion ‘propositional’ relates to ideational/content meanings, ‘textual’ pertains to cohesion making, and ‘expressive’ concerns presuppositional and other pragmatic meanings (Traugott 1995). The idea is that “historically in many cases a lexical item that originated in the ideational component later developed polysemies in [...] the textual and interpersonal domains.”

Thus, these layers are not only crucial in understanding subjectivity and subjectification, but also of utmost importance in the lexicalization-grammaticalization-pragmaticalization interface (see Section 2.2.4). Lexicalization, grammaticalization and pragmaticalization apply to these different layers. Roughly speaking, lexicalization concerns elements at the ideational level, grammaticalization is a shift from the ideational to the textual level (or within the textual level), and pragmaticalization is a shift from either the ideational or textual level to the interpersonal level. These different types of language change may be accompanied by (inter)subjectification

Traugott identified three tendencies in semantic change that are linked to the cline in (73) above. Tendency I, in (74), can feed Tendency II, in (75), and either Tendency I or II can feed Tendency III, in (76) below.

(74) **Tendency I**

meanings based in the external described situation

>

meanings based in the internal (evaluative/perceptual/cognitive) described situation

Tendency I includes a wide range of metaphorical changes from concrete to abstract notions (e.g. *to grasp*) and so-called ‘lexical subjectification’, i.e. amelioration (the development of positive meanings/stance) and pejoration (the development of negative meanings/stance). An example is English *boor* which used to mean ‘farmer’, but now denotes ‘crude person.’ A similar development applies to Dutch *boer* ‘farmer’ which also came to denote ‘crude person.’

(75) **Tendency II**

meanings based in the external or internal described situation

>

meanings based in the textual and metalinguistic situation

Tendency II includes developments leading to elements coding textual cohesion (Traugott 1995:35). In Traugott’s terminology, ‘textual situation’ refers to the situation of text-construction and ‘metalinguistic situation’ concerns the situation of performing a linguistic act. An example of the former is the development of the connective *while*, from Old English *þa hwile þe* ‘the time that’ (coding an external described situation) > *while* ‘during’ (coding the textual situation). An example of the latter is the verb *observe* that developed speech-act verb uses in the sense ‘state

that’ (coding the metalinguistic situation) from a mental-verb meaning ‘perceive (that)’ (coding an internal described situation).

(76) **Tendency III**

meanings tend to become increasingly based in the speaker’s subjective belief state / attitude toward the proposition

Tendency III is subjectification. It subsumes, amongst others, the shift from temporal to concessive *while*, the development of *be going to* into a marker of immediate, planned future (cf. example (39)) and the development of epistemic meanings (Traugott 1995:35).

In later work, Traugott revised the cline ‘propositional > textual > expressive’ once more by replacing the layers by the terms ‘nonsubjective’, ‘subjective’ and ‘intersubjective’. This cline has correlations with other tendencies in semantic change, as illustrated by Traugott & Dasher’s (2002:40) “correlated paths of directionality in semantic change,” in (77).

(77)	truth-conditional	>	non-truth-conditional
	content	> content/procedural	> procedural
	scope within proposition	> scope over proposition	> scope over discourse
	nonsubjective	> subjective	> intersubjective

De Smet & Verstraete (2006) distinguish between two types of (semantic) subjectivity in which different aspects of speaker involvement are identified. These two types are ‘ideational subjectivity’ and ‘interpersonal subjectivity’ (cf. Halliday & Hasan 1976; Halliday 1994) which, as their names suggest, relate to different linguistic components (cf. (72)).<sup>32</sup> Examples of ideational subjectivity are nouns or evaluative adjectives. Instances of interpersonal subjectivity are deontic/epistemic auxiliaries, discourse markers, illocutionary speech acts, modal adverbs or intensifying use of adjectives. Ideational and interpersonal subjectivity can be distinguished on syntactic and semantic grounds, the main difference being whether a “particular item functions in the build-up of the content of a message, or in the enactment of the speaker’s position with respect to this content” (De Smet & Verstraete 2006).

Ideational and interpersonal subjectivity show divergent syntactic behavior because linguistic items that function at the ‘interpersonal’ level cannot be subject to operations that apply to the ‘ideational’ level, such as focus, negation and *wh*-interrogation. De Smet & Verstraete use these properties as syntactic criteria in order to discriminate between ideational and interpersonal subjectivity.<sup>33</sup> They illustrate their idea by means of two case studies. The first

<sup>32</sup> De Smet & Verstraete leave out the textual level, which seems to be incorporated in the ideational level.

<sup>33</sup> Note that the tests they provide are partly overlapping with the focus and addressability tests to determine grammatical status, as presented in Boye & Harder (2012:14-15). These tests work to single out grammatical expressions and to distinguish between ideational and interpersonal subjectification because the tests apply to the content of the message. Therefore, they just show that neither grammatical items nor attitudinal/interpersonal

case study is concerned with the Dutch attitudinal adjectives *leuk* ‘nice, pleasant’ and *dom* ‘unintelligent and bloody, cursed.’ The second case study is about causal conjunctions in English, i.e. *after*, *as*, *since*, *because* and *for*.

With respect to attitudinal adjectives it is shown that *leuk* and *dom* are equally subjective in an ideational sense because they describe attitudes internal to the speaker, but *dom* in the sense of ‘bloody, cursed’ has moved to the interpersonal level. Hence, attitudes may figure either in the content of the utterance (ideational level) or in the way the speaker takes an interactive position with respect to that content (interpersonal level). The adjectives *leuk* and *dom* meaning ‘unintelligent’ have a subcategorizing function, whereas *dom* in the sense of ‘bloody, cursed’ is a stance marking device. Syntactic operations such as focus, interrogation, and negation all serve to select some aspect of content, either by highlighting it or denying it, and therefore do not apply to speaker’s attitudes that are situated at the interpersonal level (ibid:386).

As regards the causal conjunctions *after*, *as*, *since*, *because* and *for* the main observation is that they consist of two types. On the one hand *as*, *since* and *for* group together and on the other hand *because* and *after* form a subgroup. These causal conjunctions are all equally subjective in an ideational sense because they can denote causal relations, but in complex sentences, *as*, *since*, and *for* may also function at the interpersonal level. The causal conjunctions *because* and *after* allow integration into the proposition of main clauses, whereas *as*, *since* and *for* clauses do not. The importance of the speaker in *as*, *since*, and *for* clauses, and their independence from the main clause, is reflected by the fact that these clauses cannot be focused, questioned or fall within the scope of negation within the larger complex sentence. Causal clauses headed by *as*, *since* and *for* are “interactively independent units”, and as such not part of a complex proposition.

The difference between ideational and interpersonal subjectification, then, essentially boils down to the relation between ‘the speaker’s believe/stance’ and the ‘proposition.’ As Visconti (forthc.), building on Doherty (1987), notes “[p]ropositional meaning is that part of sentence meaning that is evaluable, i.e. which can, or rather must, be evaluated; non-propositional meaning is the evaluating part” and may be of an epistemic (e.g. *probably*), intentional (e.g. imperative, optative), or emotional nature (e.g. *unfortunately*).” In a recent proposal, Visconti (forthc.) suggest that an item undergoing subjectification shifts status from being an *operandum* at the propositional level of a linguistic representation to being an *operator* at the attitudinal level. In her proposal the concept of subjectification is narrowed down to ‘attitudinal subjectification’, excluding ‘lexical subjectification’ (amelioration and pejoration) and (meta-)textual subjectification (the development of ‘cohesion-coding devices’ that create text and structure).

In order to arrive at a unified account of subjectification and types of language change, I think it is necessary to retain the textual level. Therefore, I propose a modified cline on the basis

---

comments contribute to the content of the message, because they are secondary to the content and a comment towards the content respectively.

of the original insights by Halliday & Hassan (1976) and Traugott (1982; 1995), as in (78) below.

$$(78) \quad \begin{array}{ccccc} \text{ideational} & > & ((\text{textual}) & > & (\text{interpersonal})) \\ [ & \text{propositional} & ] & & [ \text{extra-propositional} ] \end{array}$$

The ideational and textual level are subtypes of the propositional level, which is composed of lexical and grammatical items that are syntagmatically related. As such lexical items may have primary status, grammatical items have secondary status (cf. Boye & Harder 2012). The interpersonal level concerns extra-propositional elements like various types of discourse markers.

Subjectification consists of two subtypes, viz. subjectification and intersubjectification, and may apply to all layers. It may, but need not, accompany processes of language change. That is, subjectification may also affect ordinary lexical and grammatical items without being involved in lexicalization, grammaticalization or pragmaticalization. Hence, both lexical(ized) and grammatical(ized) items may have subjective and/or intersubjective meanings.

In the present study (inter)subjectification is defined as in (79). The different types of subjectification correspond to a large extent to Traugott's tendencies in semantic change.

- (79) Subjectification and intersubjectification are metonymic types of semantic change that lead to increased speaker-perspective, attitude or judgment (subjectification) and attention to speaker-addressee interaction (intersubjectification). There are two subtypes of subjectification that may affect linguistic items at different layers of language, as listed in Table 9.

<b><u>(Inter)Subjectification</u></b>
<b>I. subjectification</b> [speaker perspective, attitude and judgment]
<p>- <b>ideational level</b> [connotation: amelioration and pejoration]  e.g. <i>boor</i> 'farmer' &gt; <i>boor</i> 'crude person'</p> <p>- <b>ideational level</b> [meta-linguistic meanings]  e.g. <i>observe</i> 'perceive (that)' &gt; <i>observe</i> 'state that'</p> <p>- <b>textual level</b> [text-construction]  e.g. connective <i>while</i> [<i>þa hwile þe</i> 'the time that' &gt; <i>while</i> 'during']</p> <p>- <b>textual level</b> [meta-linguistic meanings]  e.g. modal auxiliaries <i>can</i>, <i>may</i>, <i>must</i> etc.</p> <p>- <b>interpersonal level</b>  e.g. discourse markers <i>I mean</i>, <i>I think</i> etc.</p>

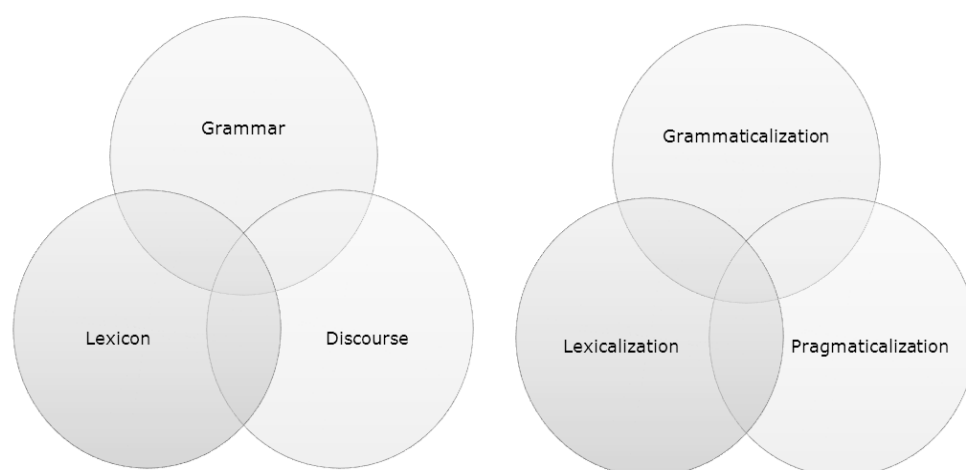
II. intersubjectification [interaction with interlocutor]
<p>- <b>ideational level</b> e.g. conversational routines, <i>thanks</i>, <i>goodbye</i>, <i>please</i> etc.</p> <p>- <b>textual level</b> e.g. modal particles <i>vel</i>, <i>jo</i>, <i>nok</i>, <i>da</i> etc. in Norwegian</p> <p>- <b>interpersonal level</b> e.g. discourse markers <i>you know</i>/<i>y'know</i>, <i>look</i>-forms, <i>well</i> etc.</p>

**Table 9.** Types of subjectification.

Both (inter)subjectification and the different types of language change relate to different layers of language (i.e. the ideational, textual and interpersonal level). This means that they may, but need not, co-occur. Like ordinary lexical items, lexicalized items may be subject to amelioration or pejoration. Clippings, such as *ism*, *ology* and *ization* may have a negative connotation. Grammaticalization may be accompanied by (inter)subjectification, but pragmaticalization necessarily involves subjectification and intersubjectification.

## 2.2 Interfaces

Both synchronically and diachronically, linguistic items may be unique, hybrid, or tripartite, with respect to their linguistic status (i.e. lexical item, grammatical item, discourse marker), domain (i.e. lexicon, grammar, discourse) as well as their diachronic development (i.e. lexicalization, grammaticalization, pragmaticalization), as illustrated in Figure 3.



**Figure 3.** Synchronic and diachronic interfaces.

In this section, some ‘problematic’, non-classifiable linguistic expressions that have repeatedly been mentioned in the literature will be discussed with regard to the lexicalization-grammaticalization-pragmaticalization interface. The converging and diverging properties of these different types of language change are summarized in Table 10 which highlights the interface areas between lexicalization, grammaticalization and pragmaticalization.

### 2.2.1 The lexicalization-grammaticalization interface

The classification of linguistic items as either lexical or grammatical, and hence their historical development as either lexicalization or grammaticalization might seem a straightforward task when taking the definitions in (58) and (45) as a point of departure. However, there are many instances of linguistic items that, at a superficial glance, can be assigned either grammatical or lexical status, but at a closer look turn out to be problematic because they are situated at the interface of lexicalization and grammaticalization.

One such an example is the temporal adverb ‘today’ in the Germanic languages. The status of this adverb is a good illustration of the confusion that may arise when assigning lexicalized or grammaticalized status to a linguistic item because *today* has been claimed to be a prototypical instance of lexicalization, but also of grammaticalization. Fischer (2007) claims that *today* is an adverb of “which no one would dispute *is* a case of lexicalization.” Similarly, Hopper & Traugott (2003:24) argue that German *heute*<sup>34</sup> ‘today’ “might be more appropriately thought of as illustrating the emergence of a new lexical item rather than of a grammatical formative.” Meillet (1958 [1912]), on the other hand, mentions *heute* as an example of grammaticalization. Brinton & Traugott (2005:1) note that “*today* is not clearly a lexical or a grammatical form, having partially concrete and partially abstract meaning.”

Two things are at issue here. First, the problem of whether *today* expresses referential meaning, relational meaning or a combination of both, and secondly, the status of (temporal) adverbs as a lexical or grammatical category. Confusion with regard to the status of *today* arises probably because this adverb is still a transparent item (*today*<sup>35</sup> < OE *to*+ *dæge* ‘at day-DAT’), unlike *heute*.<sup>36</sup> Other temporal adverbs like *tomorrow*<sup>37</sup> or *yesterday*<sup>38</sup> have become opaque

<sup>34</sup> German *heute* goes presumably back to Proto-Germanic *\*hiu dagu* ‘on this day’, which is composed of the instrumentalis form of the noun ‘day’ and a demonstrative pronoun that is only retained in fossilized expressions such as *heute* and *heuer* (< Old High German *hiuro* (< *\*hiu jāru*) ‘(in) this year’ (EWS).

<sup>35</sup> A similar formation is Dutch *vandaag* which is a contraction of the Middle Dutch dative construction *vandendaghe* (< *van [den/desen/etc.] daghe* ‘on this day’). Comparable formations are *vanmorgen/vanochtend* ‘this morning’, *vanmiddag* ‘this afternoon’ and *vanavond* ‘this evening, tonight’ (Phillippa et al. 2011).

<sup>36</sup> Note that there is a correspondence between German *heute* and Dutch *heden* ‘nowadays’ (Old Dutch *hiudo*; Middle Dutch *heden*) and the derivative form *huidig* ‘present-day’ (< (Eastern) Middle Dutch/Middle Low German *huden*; possibly formed after analogy with German *heutig*), (Phillippa et al. 2011). German *heute* can mean both ‘today’ and ‘nowadays’. German *heutzutage* and Dutch *heden ten dage* or *vandaag de dag* are mixed forms meaning ‘nowadays.’

<sup>37</sup> English *tomorrow* goes back to Middle English *to morȝen*, *to morwen*. When the final *-n*, and *e* were lost, *w* was vocalized to *-ow*, as in *arrow*, *borrow*, *sorrow*. The form *morrow* goes back to the noun *morn*. In Northern English the form *to-morn* (< *tó* + *morgenne*, dative of *morgen*, *mērgen*; *morn*<sub>N</sub> ‘beginning of the day; dawn, sunrise’) is more common (OED, Phillapa et al. 2011).



(=loss of both semantic and morphological compositionality) and because of that they are easier termed lexicalizations. One commonality is that they all belong to a small class of indexical adverbs.

In general, it is very hard to assign either lexical or grammatical status to adverbs. In this study, therefore, the notion of ‘adverb’ is a cover term that comprises subsets of lexical, grammatical and ‘mixed’ or ‘indeterminate’ (=forms that have partly concrete and partly abstract meaning) adverbs (see Chapter 4 for more details). In line with Brinton & Traugott I will argue that *today*, as well as *yesterday* and *tomorrow*, has partial referential and partial relational meaning. The meaning of temporal adverbs is partially concrete (their meanings can be rephrased as ‘on this day’, ‘on the day after today’, and ‘on the day before today’) and partially abstract (denoting time reference viewed from the speaker’s here-and-now). Hence, the rise of temporal adverbs concerns developments at the lexicalization-grammaticalization interface.

Besides the lexical or grammatical status of a linguistic item, confusion arises because of properties that lexicalization and grammaticalization have in common, such as fusion and coalescence, which lead to a loss of compositionality. For this reason the same data has been classified as both lexicalization and grammaticalization (e.g. epistemic adverbs of the MAYBE-type, see Chapter 4). According to Brinton & Traugott, only instances of fusion and phonological reduction that yield functional closed classes are cases of grammaticalization, whereas those that result in open class items may be considered lexicalizations.

A particular problematic area at the lexicalization-grammaticalization interface concerns the status of derivation. Derivation is generally associated with lexicalization because it is more idiosyncratic, less likely to be obligatory, and less productive compared to inflection, which is usually linked to grammaticalization. What causes confusion, then, is that there are two types of derivational affixes. First, there are semantic or lexical derivational morphemes like *un-* or *-ship* as in *unpleasant* or *membership*. Second, there are syntactic or grammatical derivational morphemes that are markers of syntactic recategorization, i.e. they change the category of a linguistic item. An example is agentive *-er*, as in *golfer* or adverbs formed with *-ly*, like *friendly*, *obviously*. Hence, lexicalization may result in semi-productive forms, such as restricted derivational morphemes, while grammaticalization results in forms that serve as default affixes such as inflections, but there are intermediate forms at the edge of derivation and inflection, like grammatical derivational morphemes.

Other borderline cases and problematic areas include the status of present participles (V-*ing* forms e.g. *disturbing*, *concerning*), multi-word or phrasal verbs (e.g. *look up*, *come across*), composite predicates (e.g. *have a try*, *lose sight of*), and (phrasal) discourse markers (e.g. *I think*, *anyway*, see Section 2.2.2). See Brinton & Traugott’s case studies (2005:111-140) for a thorough discussion of these problematic cases.

---

<sup>38</sup> The form *yesterday* goes back to Old English *geostran*, *gystran dæg*. Other Germanic languages employ the simple word in the sense of ‘yesterday’ (e.g. Dutch *gisteren*, but there is also the expression *de dag van gisteren* ‘the day of yesterday’; German *gestern*, Frisian *jister*), (OED, Phillipa et al. 2011).

The fact that there are a vast number of linguistic expressions at the lexicalization-grammaticalization interface goes to show that rigid distinctions between lexicon/lexicalization and grammar/grammaticalization cannot be maintained. Instead, a gradient or continuum view of the relation between grammar and ‘lexicon’ is a more appropriate representation of the actual situation. That is, the ‘inventory’ forms a continuum of lexical and grammatical items, see Section 2.1.2.

### 2.2.2 The grammaticalization-pragmaticalization interface

The classification of linguistic items as being essentially grammatical or communicative is not always easy to make, as there seem to be many instances of linguistic items that are situated at the interface of grammar and discourse. Another delicate matter is that not all linguists are willing to assign a special status to the development of discourse markers, and hence deny the existence of pragmaticalization.

The notion of pragmaticalization has been criticized for various reasons. The most frequently raised objection to pragmaticalization is that the distinction between grammaticalization and pragmaticalization is too minimal in order to speak of two distinct processes. This view is advocated by scholars who claim that grammar/grammaticalization has to be defined broadly enough in order to encompass the development of discourse markers (Brinton & Traugott 2005; Diewald 2011). In this line of thought pragmaticalization is reduced to a subtype of grammaticalization, which, according to Diewald (2011:365), “displays essential core features of grammaticalization processes, but is distinguished from other subtypes of grammaticalization processes by specific characteristic traits (concerning function and domain as well as syntactic integration).” In a similar vein, Wisher (2000:356) distinguishes between grammaticalization I and grammaticalization II. Grammaticalization I operates on the propositional level and “refers to the transformation of free syntactic units into highly constrained grammatical morphemes.” The second subtype, grammaticalization II, “operates on the textual or discourse level and concerns the development of textual or discourse markers.”

On the other hand, there are also scholars who think that a broad conception of grammaticalization so as to include discourse-pragmatic developments makes it too heterogeneous a notion that reduces its descriptive power. On this view, “movement towards discourse is not grammaticalization”, but ‘discoursivization’, i.e. “the diachronic process that ends in discourse” (Ocampo 2006), and hence the notions of discourse and grammar are best kept apart (Norde 2009:23). Note that the ‘specific characteristic traits,’ as Diewald aptly puts it, are precisely the key-defining properties of pragmaticalization as described in subsection 2.1.3, namely (pragmatic/interpersonal function, domain of discourse, syntactic externalization). An argument in favor of the distinction between grammaticalization and pragmaticalization is that the development of discourse markers runs counter Givón’s (1979:209) cline of grammaticalization, as shown in (80).

(80) discourse > syntax > morphology > morphophonemics > zero

The co-existence of two diverging orientations towards pragmaticalization leads to different classifications for one and the same item. An example is the parenthetical comment clause *I think*. As stated by Brinton & Traugott (2005:140) “the development of discourse markers, both phrasal and non-phrasal, is best understood as a process of grammaticalization.” But for Aijmer (1997:2-3) *I think* is a typical case of pragmaticalization because it involves “speaker’s attitude to the hearer.”

Modal particles (e.g. German *denn*, Dutch *ook* Norwegian *nok*), too, are linguistic items at the grammaticalization-pragmaticalization interface. Their development has been claimed to be an instance of ‘grammaticization’ (Abraham 1991), or grammaticalization accompanied by subjectification (Diewald 2011). Since modal particles are often included in the class of discourse markers (e.g. Aijmer & Simon-Vandenberg 2009), their development could also be a case of pragmaticalization (Aijmer 1997).

Detges & Waltereit (2009) see discourse markers and modal particles as two types (i.e. major classes) of a larger category of ‘pragmatic particles’, a term which they do not discuss or define any further. According to them, discourse markers and modal particles can be distinguished on functional grounds in that discourse markers “derive from strategies related to the joint coordination of interaction, while modal particles derive from strategies which refer to the status of a given proposition to the ongoing discourse” (as summarized by Beeching 2009:101). The differences between discourse markers, (81a), and modal particles, (81b), is described by Detges & Waltereit (2009:44-45) in (81) below.

- (81) a. [A] typical discourse marker signals a two-place relation which concerns the structure or form of discourse. [T]he discourse units linked by discourse markers are not necessarily identical with grammatical units at the sentence-constituent level (Hansen 1998a) In other words, the placement of discourse markers is not subject to constraints of grammatical nature. [T]heir position reflects the linear organization of discourse rather than constraints of syntactic nature.
- b. [Modal particles] are conventionally tied to particular speech-act types [and] convey one place relations with respect to illocutionary content. They are highly content-dependent and have a scope and a position which are subject to strong syntactic restrictions (see Autenrieth, 2002).

Thus, the decision of considering modal particles to be discourse markers has consequences for their status as being grammaticalized or pragmaticalized. As noted by Diewald (2011:373) the issue of pragmaticalization “[u]sually becomes relevant in certain instances of change, where the borderline between subjectification and grammaticalization is problematic, and the target, the endpoint of change in question, does not fall into the range of grammatical categories in the traditional understanding of the term.”

It is generally acknowledged by scholars with a broad conception of grammaticalization that discourse markers are in conflict with ‘standard’ cases of grammaticalization, it is, however, not made explicit why, and on what grounds, discourse should be part of grammar, besides notes on the properties that the development of discourse markers has in common with grammaticalization. Brinton & Traugott (2005:139) write “[t]hough discourse markers have primarily pragmatic meaning and carry scope over more than the sentence, they are indubitably ‘part of the grammar’, or part of the structure of the sentence (i.e. they are not extra-grammatical).”

A similar view is voiced by Traugott & Dasher (2002:159) on the status of pragmatic and procedural material that occurs outside the clause, i.e. expressive markers of various types on the left margin of the clause. Even though most material in this position is pragmatic and procedural, “it unquestionably belongs to syntax and grammar.” Since the prime concern of Traugott & Dasher is semantics, and not syntax, they leave open the question whether there is a syntactic category of D[iscourse]M[arker]s, or whether they belong to a larger set of discourse connectives that occur in the syntactic position in question.

The idea of including discourse functions in grammar is based on the observation “that many core grammatical (functional) categories such as tense, aspect and mood convey non-truth-conditional meaning (consider the pragmatic meaning of the past tense in *What was your name?* or the progressive in *Are you wanting to go now?*) [...] Brinton & Traugott 2005:129).” Although it is true that these core grammatical categories may convey pragmatic dimensions, I think these are mere instances of pragmatic use of grammatical markers, and not pragmatic meaning of these grammatical markers. Generally, these pragmatic uses are only secondary to the core grammatical meaning.

Diewald’s argumentation for classifying pragmaticalization as a subtype of grammaticalization is based on the development of the class of modal particles in German. She describes the function of modal particles as “marking a turn as noninitial by relating it to a presupposed pragmatically given unit.” This relational structure is schematized in (82).

- (82) pragmatically given unit ← (modal particle & utterance in the scope of the modal particle)

Her understanding of pragmatic meaning as a kind of relational (and hence grammatical) meaning is strongly linked to deixis,<sup>39</sup> as the following citation in (83) makes clear.

- (83) If we choose to call this a pragmatic meaning [i.e. marking a turn as noninitial by relating it to a presupposed pragmatically given unit], then, we should also call the meaning of a tense category a pragmatic meaning because it is deictic and links the linguistic level to the communicative level in exactly the same way as does the class of modal particles.

---

<sup>39</sup> Deixis is a term used “to refer to those linguistic elements which make interpretable reference only by virtue of an indexical connection to some aspect of the speech event. Typical exemplars for English include *here-there*, *this-that*, *now-then*, and *I-you*” (Sidnell 2009:114).

[...] If we chose to do that, and if we chose to call the development of these categories pragmaticalization, then we would have to call the development of all deictic grammatical categories pragmaticalization, because there is no definitional difference between them and the development of discourse markers (Diewald 2011: 382-3).

On this view, “the only difference between them [deictic grammatical categories and discourse markers KB] is their respective formal realization and their specific semantic/functional domain” (ibid), as they have some kind of relational meaning as a common denominator, which allows them to be subtypes of the same superordinate process. In my view, however, the different formal realization and semantic/functional domains of deictic grammatical categories and discourse markers are exactly their distinctive properties. Moreover, though textual subjectification can be seen as a kind of deixis, subjectification in general cannot be reduced to deixis.

In sum, the fact that there are a vast number of linguistic expressions at the grammaticalization-pragmaticalization interface goes to show that rigid distinctions between grammar/grammaticalization and discourse/pragmaticalization cannot be maintained. Instead, a gradient or continuum view of the relation between grammar and discourse is a more accurate representation of the actual situation. Moreover, it is important to keep in mind that there is a difference between grammaticalization accompanied by (inter)subjectification, as in the case of modal particles, and pragmaticalization which is composed of (inter)subjectification and syntactic externalization, as is the case for discourse markers such as *you know*, *I mean*.

### 2.2.3 The pragmaticalization-lexicalization interface

Confusion about lexical or pragmatic status concerns primarily routinized items and expressions, and discourse markers. Some discourse markers have been subsumed under the heading of both lexicalization and pragmaticalization, which suggest the existence of certain parallels between lexicalization and pragmaticalization. An example of a linguistic expression that has been assigned widely diverging statuses is *I think*.

The phrasal discourse marker *I think* has, because it is fixed and partially fused (Brinton & Traugott 2005:137), been claimed to be an instance of grammaticalization (Thompson & Mulac 1991; Brinton 1996; Brinton & Traugott 2005, see Section 2.2.2). It has been labeled pragmaticalization as it “permits, for example, extensions of meaning involving the speaker’s attitudes to the hearer or to the message” (Aijmer 1997:3), and it has been conceived of as a case of lexicalization (Schiffrin 1987:319; Fischer 2007). Fischer (2007:112), cited in (84), concludes that:

- (84) parenthetical phrases like *I think* etc. are best seen as formulaic tokens, undergoing lexicalization rather than grammaticalization. In this process, they lose some referential content, being narrowed down to a more epistemic, evaluative meaning. In non-standardized languages they are likely to form one lexical unit in the course of time [...].

This view is in conflict with the definition of lexicalization in the present study because it states that the loss of some referential content and epistemic, evaluative meaning are signs of a lexicalization process, whereas lexicalization in the present study is characterized by reinterpretation of referential meanings. A second argument against treating discourse markers as being lexicalized is that they do not belong to a major category, nor do they convey the main point in a linguistic message.

Another debatable area concerns the status and development of so-called conversational routines, such as *thank you, please, I'm sorry, goodbye* etc. Blank (2001:1604) considers the transformation of a speech-act into a word, for example *goodbye* (< *God be with you*) as an instance of lexicalization. Also Aijmer (1997) sees such 'conversational routines' as lexicalizations. The question is whether these expressions are lexicalized formula's or a particular type of pragmatic marker. An argument against lexical status is that they do not belong to a major category and do not convey referential meaning.

One more borderline case is English *ish* 'sort of'. The form *ish* is both a verb (an echoic nonce-word, meaning to make the sound *ish* or *sh*) and an adverb that derives from the suffix – *ish*. As an adverb its function is "qualifying a previous statement or description", which makes it a "conversational rejoinder", meaning 'almost, in a way, partially, vaguely' (OED). Consider the examples in (85) below.

- (85) a. *One of those neatly crafted middle-brow plays which, because they have a pleasantly happy ending (well, **ish**), might make people think that they've been handed a soft option* (OED, Sunday Times).
- b. *Mr. Langmead, speaking by telephone from London, hesitated. 'Ish,' he said, employing the international shorthand for slight hedge* (OED, N.Y. Times).

An argument in favor of *ish* being lexicalized is that it is a clipping, the meaning of which is not completely transparent. On the other hand, *ish* conveys pragmatic meanings and occurs sentence-externally, which pleads for pragmaticalized status.

In sum, there is some overlap between lexicalization and pragmaticalization. They are, however, more clearly distinguishable on categorical, semantic and functional grounds than for example lexicalization and grammaticalization, or grammaticalization and pragmaticalization.

#### 2.2.4 The lexicalization-grammaticalization-pragmaticalization interface

As described in the previous subsections, lexicalization, grammaticalization and pragmaticalization have much ground in common, but there are also fundamental differences that can be traced back to the different functional domains to which these processes apply. In order to distinguish between lexicalization, grammaticalization and pragmaticalization it is important to distinguish between different domains of language: lexicon (ideational level), grammar (textual level) and discourse (interpersonal level), cf. the synchronic and diachronic interfaces in Figure

3. Not surprisingly, then, lexicalization operates at the ideational level, grammaticalization applies to the textual level (a shift from the ideational to textual level, or a shift within the textual level), and pragmaticalization operates at the interpersonal level (a shift from ideational or textual level to interpersonal level). The different types of language change may be accompanied by (inter)subjectification. In this study, lexicon, grammar and discourse are subcomponents of a larger ‘inventory’ of linguistic expressions. They form a continuum which is reflected by the fact that some linguistic classes may contain lexical, grammatical and communicative subsets e.g. adverbs (e.g. manner adverbs, sentence adverbs, speech-act adverbs respectively). The converging and diverging properties of lexicalization, grammaticalization and pragmaticalization are summarized in Table 10 below. Again, the sign ‘+’ stands for a key-defining property, ‘-’ denotes that a certain feature does not apply, and ‘(+)’ represents characteristics that may, but need not, be involved in a certain type of language change.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	± <sup>40</sup>
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis*</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-
<b>reinterpretation</b>					
<b>-metaphor/metonymy<sup>41</sup></b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

<sup>40</sup> In case one considers the whole class of discourse markers to be a ‘minor category’ (e.g. Dostie 2009) categorical reanalysis would also be a key-defining property of pragmaticalization. However, since discourse markers may have both lexical and grammatical sources, which do not necessarily change category when becoming a discourse marker, category status (major/minor) does not matter. Extra-propositional status, and hence no grammatical connection to the proposition, of an expression is the key-defining property in pragmaticalization.

<sup>41</sup> Brinton & Traugott (2005:107) note that metonymization and metaphorization may be appropriate for both lexicalization and grammaticalization, but they do not necessarily refer to the same types of metonymy or metaphor. Metonymy in lexicalization is more likely to be driven by social custom and encyclopedic knowledge (e.g. *forget-me-not*). In grammaticalization it is driven by more strictly linguistic meaning (e.g. *be going to*).

<b>ii. Primitive changes</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>-phonology/phonetics</b>					
° loss of phonological/phonetic substance	(+)	-	(+)	(+)	(+)
<b>-morphology</b>					
° loss of morphological compositionality	+	-	(+)	(+)	(+)
° loss of morphosyntactic properties	-	-	+	(+)	(+)
<b>-syntax</b>					
° loss of syntactic variability	-	-	+	+	-
° loss of syntactic autonomy	-	-	+	+	-
<b>-semantics</b>					
° loss of semantic substance	-	-	+	+	+
° loss of semantic compositionality	+	-	(+)	(+)	(+)
<b>-discourse/pragmatics</b>					
° subjectification	(+)	(+)	(+)	(+)	+
° intersubjectification	(+)	(+)	(+)	(+)	+
<b>iii. Side effects of change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
-paradigmaticization	-	-	+	+	(+)
-obligatorification	-	-	(+)	(+) <sup>42</sup>	-
-condensation	-	-	+	+	-
-layering/divergence/specialization/persistence	+	+	+	+	+
-productivity	-	+	+	+	+
-frequency	-	+	+	+	+
-typological generality	-	(+)	+	(+)	(+)

**Table 10.** Converging and diverging properties of lexicalization, grammaticalization and pragmaticalization.

In the present study, the essence of lexicalization, grammaticalization and pragmaticalization is defined in terms of formal reanalysis and semantic reinterpretation within or between different domains of language (i.e. lexicon, grammar, discourse). Lexicalization, grammaticalization and pragmaticalization are conceived of as composite changes that are made up of basic mechanisms of change, a subset of correlated primitive changes at different linguistic levels (phonology/phonetics, morphology, syntax, semantics and discourse/pragmatics) and the concomitants of these (micro-)changes.

In the literature there is a tendency to pick out a subset of properties, on the basis of which one assigns the status of lexicalization, grammaticalization or pragmaticalization to a particular change. For example, reported instances of grammaticalization are hardly ever characterized by all characteristics of grammaticalization, but an arbitrary subset of features is generally considered to be sufficient to call a change an instance of grammaticalization. Not only

<sup>42</sup> Secondary grammaticalization leads to grammatically obligatory items in case of inflectional properties. Other types of grammatical markers/expression are not always obligatory, e.g. (modal) auxiliaries.



do predefined lists of properties run the risk of being circular (e.g. grammaticalization consists of the properties A, B and C. Hence, property A+B+C is a case of grammaticalization), they may also lead to misconceptions and mismatches. In addition, most characteristics that have been mentioned in the literature are not properties of the change proper but mere signs of ongoing change, i.e. side effects. In the present study a distinction is made between properties of the change proper (reanalysis and reinterpretation) accompanying primitive changes, and the concomitants of these (micro-)changes. These side effects are signs of ongoing change and may in turn be used as a diagnostic to identify potential cases of lexicalization, grammaticalization or pragmaticalization.

The labels of grammaticalization, lexicalization and pragmaticalization, as they are currently defined in the literature, seem to be too diverse, which makes these notions heterogeneous concepts that capture widely diverging developments. Part of this variation is due to the fact that, for example, not all grammatical items share the same properties. Modal auxiliaries and inflections are both grammatical elements, but they differ in properties such as subjectification and bonding. Because of this, it seems impossible to characterize changes uniquely as an instance of grammaticalization, since individual cases differ in possible accompanying changes and side effects. Moreover, since there is no consensus on definitions of the different X-izations one might wonder whether a certain change (e.g. grammaticalization) in view A, would also qualify as an instance of grammaticalization from perspective B.

As has become clear from defining lexicalization, grammaticalization, pragmaticalization and their interfaces, categorization is a useful theoretical construct to create order out of inconvenient arrangements of abstract notions. In practice, however, (predefined) categories often turn out to be inapplicable to concrete instances of change. The next chapters deal with case studies of linguistic items that defy strict categorization. In order to be able to analyze and describe these data, I propose to let go of predefined categories and instead use a clustering approach. The advantage of clustering, as opposed to categorization, is that one can identify both prototypical and marginal instances of a certain type of change. Moreover, a clustering approach can deal with the gradient nature of lexicalization, grammaticalization and pragmaticalization, and captures borderline cases and interface areas.

## Chapter 3

### 3. Case study I: modal auxiliaries MUST/MAY

#### 3.1 Introduction

This chapter presents a comparative synchronic corpus investigation of the formal and semantic properties of the Mainland Scandinavian modals ‘must, may’, i.e. Danish *måtte*, Norwegian *måtte* and Swedish *må*, *måtte* and *måste*. These modals will be referred to as Mainland Scandinavian MUST/MAY whenever all of them are discussed together. The focus of this chapter is the development of modal and postmodal meanings in MUST/MAY in relation to grammaticalization (Section 2.1.1) and (inter)subjectification (Section 2.1.4). Modal meanings involve possibility and necessity as paradigmatic variants,<sup>43</sup> which is the case for the domains of dynamic, deontic and epistemic modality. Postmodal meanings, e.g. optative or concessive meanings, are meanings that originate in either possibility or necessity (van der Auwera & Plungian 1998:1).

The development of modal and postmodal meanings is a well-known instance of semantic change which is commonly accompanied by (inter)subjectification. The modals *må*, *måtte* and *måste* have developed a wide range of different meanings and uses in the domain of necessity, possibility and beyond, i.e. notions like ability, capacity, possibility, permission, obligation, uncertainty, probability, concession and wish. Moreover, modals are prototypical instances of grammaticalization, because they are forms on their way from full verb to auxiliary status. This combination of formal and semantic changes makes modals very suitable to elaborate on the relation between grammaticalization and (inter)subjectification.

Some examples of modal and postmodal meanings are given in (86). The examples in (86a-c) are taken from Van der Auwera et al. (2009), example (86d) is taken from Traugott & Dasher (2002).

- |      |    |  |            |
|------|----|--|------------|
| (86) | a. | <i>You <b>may</b> go now.</i>                                  | deontic    |
|      | b. | <i>He <b>may</b> be home, or he may not – I don’t know.</i>    | epistemic  |
|      | c. | <i><b>May</b> he live a hundred years!</i>                     | optative   |
|      | d. | <i>She <b>may</b> jog, but she sure looks unhealthy to me.</i> | concessive |

---

<sup>43</sup> That is, “as constituting a paradigm with two possible choices, possibility and necessity” (van der Auwera & Plungian 1998:80). This is a narrow definition of modality, but since Mainland Scandinavian MUST/MAY cover all kinds of meanings in the domain of possibility and necessity, this narrow definition suffices for the present study. The definition applies to grammaticalized verbal modality markers but will be valid for other types of markers, too (ibid:87).

The aim of this chapter is to find out whether the development of Mainland Scandinavian MUST/MAY is in line with cross-linguistically attested grammaticalization paths, and whether these modals differ in degrees of grammaticalization and (inter)subjectification. The following research questions, as stated in (87) below, will be addressed.

- (87) What are the differences and similarities between Danish, Norwegian and Swedish with respect to the semantic distributions and formal properties of the modals *må*, *måtte* and *måste*? More specifically, do the Mainland Scandinavian languages differ in the degree of grammaticalization and (inter)subjectification for these modals and how do grammaticalization and (inter)subjectification interact in the development of these modals?

In order to determine the degree of grammaticalization for Mainland Scandinavian MUST/MAY both formal and semantic criteria will be examined. As regards the formal side of grammaticalization, factors like the inflectional paradigm (full or defective), syntactic positions (fixed or free) and the number of specific constructions and contexts in which a form may occur (few or many) is taken into account. Defective inflectional paradigms, fixed syntactic positions and frequent occurrence in specific constructions and contexts are all signs of (advanced) grammaticalization. The degree to which MUST/MAY have lost verbal properties is essential in determining the degree of auxiliation for these modals.

With respect to the semantic side of grammaticalization, I will compare the distributions of the different meanings and uses of *må*, *måtte* and *måste* in the Mainland Scandinavian languages. More specifically, differences in proportions between premodal, modal and postmodal meanings will be examined. The idea is that premodal, or lexical source concepts, give rise to modal meanings which in turn may develop into postmodal meanings. A simplified developmental path for modal and postmodal meanings is given in (88) below.

- (88) premodal      →                      modal                      (→                      postmodal)

A widely attested tendency in the development of grammatical items is that their meanings become increasingly (inter)subjective over time (e.g. Traugott 1989; 1995; 2003). As regards (inter)subjectification, the perspective is in line with Traugott (e.g. 2010). That is, subjectification is seen as a process of semantic reanalysis through which expressions of speaker-reference or speaker-involvement arise, and intersubjectification is taken to be a process of semantic reanalysis through which expressions of speaker-writer and addressee-reader interaction develop (cf. Section 2.1.4). Recall that (inter)subjectification may affect linguistic items at different linguistic layers (ideational, textual and interpersonal level, cf. Halliday & Hassan 1976; Traugott 1982;1989;1995) and may, but need not, accompany processes of language change (e.g. grammaticalization).

It is hypothesized that, in the course of auxiliation, MUST/MAY have lost the prototypical

properties of main verbs and that their semantic development follows the well-known path from premodal to modal (and possibly postmodal meaning). It is expected that these modals have become more subjective over time because of their modal and postmodal meanings. The synchronic status, both formal and semantic, as well as the historical development of these modals will be discussed. Once the status and properties of *må*, *måtte* and *måste* have been established, their development will be analyzed with respect to characteristics of grammaticalization and subjectification to see how empirical data matches theoretical claims and observed tendencies within grammaticalization studies.

This chapter is organized as follows: Section 3.2 discusses modals in general and epistemic modals more specifically; Section 3.3 is about Mainland Scandinavian MUST/MAY and their etymology (Section 3.3.1) and general line of development (Section 3.3.2). Section 3.4 describes the sources and method used in this case study; Section 3.5 presents the results of the comparative corpus investigation of *må*, *måtte* and *måste*. In Section 3.6 the development of *må*, *måtte* and *måste* is discussed with respect to degrees of grammaticalization (3.6.1) and (inter)subjectification (3.6.2). Finally, Section 3.7 contains a summary, discussion and the conclusions of this case study.

### 3.2 Modals

The modals, also known as modal auxiliaries or modal verbs,<sup>44</sup> are undoubtedly the most intensively discussed linguistic items when it comes to studies of (epistemic) modality. Whether modals are to be regarded as full verbs or auxiliaries has been a much discussed topic in the literature. Modals are not clearly verbal or pure grammatical items. Rather, they are forms on their way from main verb to auxiliary status, passing intermediate stages with varying degrees of auxiliarity. Because of this, modals have properties of both full verbs and auxiliaries. The development of modals out of full verbs into auxiliaries is a well-known instance of grammaticalization (Hopper & Traugott (2003:55-58), which is also known as ‘auxiliation’ (Heine 1993; Kuteva 2001).

Formally, modals behave like grammatical elements as they, to greater or lesser extent, have lost the properties of main verbs. Signs of auxiliation are, inter alia, deficient verbal paradigms, loss of person, number or tense inflections, loss of transitivity, as well as fixed syntactic positions because they become dependent on the main verb.

Semantically, there are cross-linguistic diachronic and synchronic tendencies with respect to the meanings that may be expressed by modals (Bybee et al. 1994, Traugott & Dasher 2002; van der Auwera & Plungian 1998). Modals have the ability to express different aspects of the modal spectrum. One and the same modal may convey dynamic, deontic and epistemic meanings or a subset of these, cf. example (86) which illustrates some of the different meanings of *may*.

There are different degrees of grammaticalization for cognate modals in the Germanic languages (Nuyts 2001; Mortelmans et al. 2009). English has the most grammaticalized modal system, which is also identified as a separate grammatical category in the English language.

---

<sup>44</sup> The term ‘modal’ is neutral with respect to category status, unlike ‘modal auxiliary’ or ‘modal verb’ (Eide 2005).

However, the distinctive properties of English modals do not equally apply to their Germanic cognates as these (still) have properties in common with main verbs to varying extents. For example, the syntactic distribution of modals in Germanic languages other than English is less restricted. They may co-occur with other modals (though in a limited way), for example Dutch *moet kunnen* ‘must can’ or Norwegian *vil måtte* ‘will must’. They also have main verb-like inflectional paradigms (i.e. infinitival forms and inflection for person, number and tense).

But there are gaps in the paradigm, which distinguishes modals from full verbs. For instance, modals generally cannot occur in passive constructions, there are no present participle and imperative modal forms.<sup>45</sup> Thus modals share properties with main verbs (inflection), auxiliaries (infinitival complements), but some properties are unique to the modals only (the expression of various modal dimensions).

Epistemic modals form a relatively closed subset within the modal paradigm. In general, the epistemic meaning of a modal is secondary to its other usage(s) (Nuyts 2001:172). For example, ‘must’ is primarily a deontic modal, but may express epistemic modality<sup>46</sup> as well. When ‘must’ expresses epistemic modality, it often occurs in a fixed configuration like ‘must be X’ or ‘must have been X’. When co-occurring with other modals, there is a tendency for epistemic modals to occur leftmost in the clause. There is an increase in semantic scope for epistemic modals, when compared to deontic and dynamic meanings, because epistemic qualifications affect the content of the entire clause, see also the hierarchy of qualificational categories in Figure 2.

### 3.3 Mainland Scandinavian MUST/MAY

The prototypical meanings of Mainland Scandinavian MUST/MAY are given in (89) below, (cf. Teleman et al. 1999; Eide 2005:82-84). Swedish *må* and *måtte* can be seen as two different modals or as two variants of the same modal (Teleman et.al. 1999). Historically they derive from the same verb (<Old Swedish *magha*), *må* being present tense and *måtte* being past tense, but nowadays they have significantly different semantic distributions (see also Figure 5 and Figure 6 in Section 3.4). In this chapter I will treat them as two different modals because they no longer express present and past tense of the same modal.

(89)	language	modal	gloss
	Norwegian	<i>måtte</i>	must (deontic, epistemic)
	Danish	<i>måtte</i>	must, may (deontic, permissive, epistemic)
	Swedish	<i>må</i>	may, should (permissive, optative, concessive)
		<i>måtte</i>	may, must (deontic, epistemic, optative)
		<i>måste</i>	must (deontic, epistemic)

<sup>45</sup> There are some exceptions to this general tendency, but these are rare (cf. Eide 2005; Mortelmans et al. 2009).

<sup>46</sup> Nuyts (2001) classifies this inferential meaning of *must* as evidential, not epistemic. Inferential meanings are on the border between epistemic modality and evidentiality, as they include both the speaker’s likelihood estimation and the source of information. As likelihood estimations are based on what is known by the speaker, I will group inferential meanings under epistemic meanings. See also Section 1.1 on epistemic modality and related domains.

In general, the Mainland Scandinavian modals conform to the properties mentioned in Section 3.2. They are forms on their way from main verb to auxiliary status and have the ability to express various modal dimensions. Throughout the years, the modals *må*, *måtte* and *måste* have developed a wide range of modal and postmodal meanings. The semantic distributions of *må*, *måtte* and *måste* include, inter alia, deontic meanings as illustrated with Danish *må* in (90a), optative meanings exemplified with Swedish *måtte* in (90b), epistemic meanings as represented by Swedish *måste* in (90c), concessive meanings as shown by Swedish *må* in (90d) and eventuality meanings as is the case for Norwegian *måtte* in (90e).

- (90) a. *De grænser, man sætter, **må** håndhæves.* Danish  
 The borders one sets must maintain.<sub>PAS</sub> KDK  
 ‘The limits one sets have to be maintained.’
- b. ***Måtte** vädret vara cykelvänligt* Swedish  
 May.<sub>PAST</sub> weather.the be cycling.friendly SK  
*i sommar och resten av året.*  
 in summer and rest.the of year.the  
 ‘May the weather be cycle-friendly this summer and the rest of the year.’
- c. *Allt talar för det utom hennes* Swedish  
 Everything speaks for it except her SK  
*ålder, hon **måste** ha varit nästan 30.*  
 age she must have been almost 30.  
 ‘Everything points to this, except for her age, she must have been almost 30.’
- d. *Bryssel **må** vara delat, men det är* Swedish  
 Brussels may be divided but it is SK  
*franska som gäller på skolan.*  
 French that counts in school  
 ‘Brussels may be divided, but it is French that counts in school.’
- e. *De som eventuelt **måtte** tro at* Norwegian  
 Those that possibly must.<sub>PAST</sub> believe that NAK  
*hun har sikret seg økonomisk ved å*  
 she has insured herself economically by to

*komme med på LPGA-touren, tar veldig feil.*  
 come with on LPGA-tour.the, take very wrong.  
 ‘Those who might think that she insured herself financially  
 by means of participation in the LPGA-tour, are very much mistaken.’

The Swedish modals *må*, *måtte* and *måste* have defective paradigms. They lack infinitival and perfect forms<sup>47</sup> and do not distinguish between present and past tense. Danish and Norwegian *måtte* do have infinitival and perfect forms, as well as tense inflections. Unlike main verbs, Mainland Scandinavian MUST/MAY does not have present tense endings in *-er/-r*, e.g. *\*mår*,<sup>48</sup> *\*måtter* *\*måster*. There is also no agreement marking, but this is a general property of Mainland Scandinavian verbs, not a specific property of Mainland Scandinavian modals.

As opposed to main verbs, modals do not have present participles *\*mående*/*\*måttende*/*\*måstende*, imperative forms *\*må!/\*måtte!/\*måste!*, nor passive forms<sup>49</sup> *\*måttes/måstes/mås*, *\*bli/er/får måttet/måst* (cf. Eide 2005). Combinations with other core modals are possible for Danish and Norwegian *måtte*, though with restrictions: *måtte kunne*, *\*måtte burde*, *\*måtte skulle*, *måtte ville*. Swedish *må* and *måtte* cannot be combined with other modals and Swedish *måste* may only combine with *kunna* and *vilja*. The inflectional paradigms for Mainland Scandinavian MUST/MAY are given in Table 11.

	infinitive	present	past	perfect	gloss
Danish <i>måtte</i>	<i>at måtte</i>	<i>må</i>	<i>måtte</i>	<i>har måttet</i>	‘must, may’
Norwegian <i>måtte</i>	<i>å måtte</i>	<i>må</i>	<i>måtte</i>	<i>har måttet</i>	‘must’
Swedish <i>må</i>		<i>må</i>			‘must, may’
Swedish <i>måtte</i>			<i>måtte</i>		‘must, may’
Swedish <i>måste</i>	( <i>att måsta</i> ) <sup>50</sup>	<i>måste</i>	<i>måste</i>	<i>har måst/(måstat)</i> <sup>51</sup>	‘must’

**Table 11.** Inflectional paradigms for *må*, *måtte* and *måste*.

The degree to which modals have lost main verb properties indicates their auxiliariness. Danish and Norwegian *måtte* are modal auxiliaries, but they have retained properties of main verbs to various degrees. For Norwegian *måtte* there are as many as 14,010 occurrences of the infinitival form *å måtte* in *Norsk Aviskorpus*, and for Danish the infinitival form *at måtte* occurs 396 times in *KorpusDK*. For Swedish *måste* the perfect form *har måst* occurs 312 times in *Språkbankens konkordanser*, but perfect forms are more frequent in Danish, 1732 occurrences of *har måttet* in

<sup>47</sup> SAOB mentions infinitival *måtta*, and supinum *måttat* as new forms to *måtte*. However, I did not come across instances of these ‘new forms’ in the corpus.

<sup>48</sup> Swedish lexical *att må* ‘feel’ does have tense inflections: *mår* – *mådde* – *har mått* (e.g. SAOL).

<sup>49</sup> In the Scandinavian languages there are two different types of passive constructions. One is the so-called *s-passive*, the other a passive construction with either *bli* ‘become’, *være* ‘be’ or *få* ‘get’, as in *Middagen serveres klokka sju* and *Middagen blir servert klokka sju* ‘Dinner will be served at 7 o’clock’, respectively.

<sup>50</sup> The form *att måsta* is a newly formed infinitive (SAOL, SAOB).

<sup>51</sup> The perfect forms *har måstat* (SAOB) is placed within brackets because of its infrequent occurrence.

*KorpusDK*, and much more frequent in Norwegian, 11,252 hits for *har måttet* in *Norsk Aviskorpus*. This goes to show that the Swedish modals have progressed most when it comes to the loss of main verb properties.

### 3.3.1 Etymology of Mainland Scandinavian MUST/MAY

Originally, Danish *måtte*, Norwegian *måtte*, and Swedish *må*, *måtte* were lexical verbs meaning ‘to be strong, to have the power/strength’ (< Old Danish *mughu/muga*, Old Norse *mega*, Old Swedish *magha*; cf. Birkmann 1987). Swedish *må* and *måtte* are the original cognates of Danish and Norwegian *måtte* but have now been largely replaced by *måste*. Swedish *måste* is a borrowing from Middle Low German *moste* which is the imperfect tense of *moten* ‘to have permission / to be obliged’ (Wessén 1965:243).

The semantics of *må*, *måtte* and *måste* is a complex matter as notions associated with possibility and necessity and the like are closely related and overlapping meanings. The precise etymology of Germanic ‘must’ is unknown. It derives from a root *\*mōt-* but both its original meaning and exact semantic development remain subject to speculation. There are two main proposals. According to one line of thought ‘must’ has developed out of a meaning ‘can, have the opportunity.’ Another proposal is that ‘must’ originates in a meaning ‘to measure.’

The Germanic modal ‘must’ is a so-called preterite-present verb which means that it derives from a strong verb whose past tense came to be used as present tense. Subsequently, new infinitival and past tense forms had to be created. In the case of ‘must’ we get the infinitival form *\*mōtan-* and past tense form *\*mōt-ta-*, which became *\*mōssa-* because of regular sound change. However, this form could not be recognized as a past tense form. In order to mark past tense –*t* was inserted, which resulted in *\*mōsta-* as the past tense form of *\*mōtan-* (Phillips et al. 2011). Examples of this development are Dutch *moest*, German *musste* and English *must*.

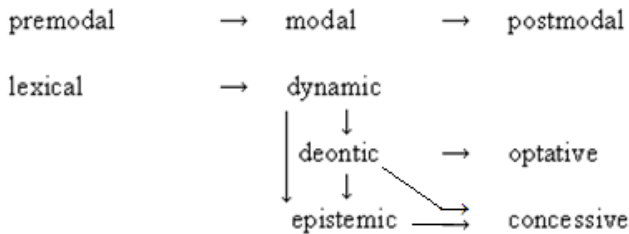
For Danish and Norwegian there is no split into different forms for ‘may’ and ‘must’ like there is in English (*may* and *must*), Dutch (*mogen* and *moeten*) or German (*mögen* and *müssen*). Danish and Norwegian *måtte*, and Swedish *må*, *måtte* ultimately derive from another preterite-present verb, ‘may’, which derives from the root *\*mag-*. The original meaning of ‘may’ is ‘be able to, can.’ Already in the oldest stages of its development a wide variety of derived meanings were available, including ‘to have the opportunity’ and ‘to have permission’ (Phillips et al. 2011).

### 3.3.2 Development of Mainland Scandinavian MUST/MAY

The development of modal and postmodal meanings is a well-known instance of semantic change that follows predictable developmental paths. That is, there are cross-linguistic synchronic and diachronic correlations between premodal, modal and postmodal meanings, as illustrated in Figure 4 below, which is a simplified generalization based on work by Bybee, Pagliuca & Perkins (1994), Van der Auwera & Plungian (1998), Heine & Kuteva (2002) and Traugott & Dasher (2002). As will become clear in the remainder of this section, there are



contiguous relations between premodal, modal and postmodal meanings as premodal meanings give rise to modal meanings which in turn may develop postmodal meanings.<sup>52</sup>



**Figure 4.** Cross-linguistic correlations for premodal, modal and postmodal meanings for MUST/MAY.

As we have seen, the predecessors of *må* and *måtte* started out as lexical verbs meaning ‘be strong, have the power/strength’. Nowadays, Swedish still has a lexical verb *att må (bra/illa)* ‘to feel (well/bad)’ which has the same origin as the modals *må* and *måtte*. This situation results from a lexical split rather than degrammaticalization, as claimed by van der Auwera & Plungian (1998:105,116). The lexical variant derives from one of the earliest meanings of *må*, and developed simultaneously with, -and not out of-, the modal meanings of *må*. Therefore, its development cannot be an instance of degrammaticalization.

The meaning ‘to feel (well/bad)’ is closely related to the original meaning ‘to be strong/to have the power’ as it arose in contexts in which one is inquiring after the power/strength or well-being of someone, as in modern Swedish *Hur mår du?* ‘How are you doing?’ (Andersson 2007:67). The original meaning ‘be strong, have the power/strength’ did not survive in Danish, Norwegian and Swedish, but is retained in verb forms like Swedish *förmå*, and Danish and Norwegian *formå*, all meaning ‘to have the ability/capacity.’

From the original meaning ‘to be strong/have the power’ it is only a small step to dynamic meanings which are concerned with abilities and capacities. Once these meanings were established, they developed further into ‘to have the opportunity’ which in turn developed further into the deontic meaning ‘to have permission to.’<sup>53</sup> The necessity meaning arose later, as is the case for German *müssen* ‘must’ which originally meant ‘to have the opportunity’ (Falk & Torp 1903).

Traugott & Dasher (2002:124) propose a plausible grammaticalization path for necessity meanings out of permission meanings in terms of scalar strengthening from relatively weak permission to deontic obligation. Contexts of denied or negative permission, i.e., ‘you may not’, may give rise to obligation meanings, i.e., ‘you must’. Since ‘you may not’ implies ‘be obliged

<sup>52</sup> This view on semantic change is in line with *The Invited Inferencing Theory of Semantic Change* as proposed by Traugott & Dasher (2002: 34-40). See Andersson (2007:166-184) for an analysis in terms of *force dynamics* (e.g. Talmy 1988, 2000; Sweetser 1990; Brandt 2004).

<sup>53</sup> See Andersson (2007) for a detailed description of the development from lexical > dynamic > deontic meanings for *må* and *måtte* in Old Swedish.

to not', the weak permission meaning is strengthened to deontic obligation. Eventually, the stronger implication of 'you may not', i.e. 'you must', became the conventionalized meaning. This is a well attested tendency as there seems to be a regular unidirectional shift among modals from e.g. "not necessary that" > "necessary that not" (cf. van der Auwera 2001). See Traugott & Dasher (2002) for a thorough discussion of the factors involved in this complex change.

For Swedish *må* the expression of necessity was fairly restricted in Old Swedish (Andersson 2007:199). This might have been a reason why Middle Low German *moste* 'to have permission/to be obliged' was borrowed to express deontic, and at a later stage, also epistemic necessity. Deontic meanings may also give rise to optative and epistemic meanings. A probable grammaticalization path for optative meanings is a development from deontic possibility (permission) to optative. The idea is that "a wish is like an appeal to circumstances (destiny) to allow the realization of a state of affairs" (van der Auwera & Plungian 1998:107).

It is generally assumed that epistemic necessity derives from deontic necessity (obligation). A clause like *He must be in the office* is ambiguous between deontic (i.e. he is obliged to be in the office) and epistemic reading (i.e. it must be the case that he is in the office). The transfer, or invited inference, of the obligation meaning is that when something is obligatory it has a high probability to occur. In his study of *må* and *kunna* in Old Swedish, Andersson (2007) proposes two plausible bridging contexts for the rise of epistemic meanings of *må* and *måtte*. The dominating meanings in Old Swedish are permission, freedom of choice and prohibition (=negated permission). Epistemic meanings occur sporadically, and only in religious texts. They are found in contexts in which evidence for a speaker-judgment or conclusion is explicitly addressed, as in example (91a). Andersson identifies and exemplifies two bridging-contexts for the rise of epistemic and epistemic-concessive meanings of *må*. He considers conditional clauses with animate subjects and a cognitive verb, as in example (91b), to be the main bridging context between root (=dynamic + deontic) and non-root (=epistemic) meanings (Andersson 2007:205).

- (91) a. *Aristodemus swarar drikker thu thz eter som the thre*  
 Aristodemus answer drink you that gift which the three  
*mæn drukko før aff ok doo: ok halder thu sidhan*  
 man drank before of and died and keep you then  
*liff oc helso: tha ma thin christus wara sander gudh.*  
 life and health then may your Christ be true God.  
 'A answers, if you drink that poison, which the three men drank of before and died, and keep alive and stay in health, then your Christ may be the true God.'
- b. *Rådhis han glödhina/ oc faklar ey fingrom i rödha*  
 fear he ember/ and put not fingers in red  
*elden tha maghin i wita at biskopen hawer sant.*  
 fire then may you know that bishop has true.

‘If he fears the ember and does not put his fingers in the red fire, then it is possible for you to know (hence conclude) that the bishop was right.’

Andersson hypothesizes that *må* in (91b) still expresses root modality. The idea is that the speaker “presents the possibility for the subject-participant (conversational partner) to realize some fact, given some external evidence. When there is no concluding participant present, the concluding becomes associated with speaker, thereby paving the way for a more speaker-oriented modality” (ibid).

The second bridging context pertains to potential epistemic-concessive readings of *må* and is found in theological-argumentative contexts in paraphrases of the Old Testament, as in (92) below.

- (92) *Nw magho män ther styggias widh/ at en hælgher*  
 Now may men be frighten at/ that a holy  
*patriarcha hafdhe fyra husfrwr oc än twa syster [...]*  
 patriarch had four wives and to that two sisters  
*Än iacob syndar ey mot natwrinne.*  
 but Jacob sin not against nature.  
 ‘Now may men be frightened at, that a holy patriarch had four wives, and to that two sisters, but Jacob do not sin against nature.’

As argued by Sweetser (1990), modals also express a kind of modality which cannot straightforwardly be identified as either root or non-root modality. According to Andersson, (92) is such a case because *må* does not express pure possibility or permission for some men to be frightened, nor does it express the speaker’s conclusion about some men’s reaction. Rather, “the reaction is presented as hypothetical and reasonable considering the words in the Bible” (ibid:206).

Concessive meanings come in various types and are likely to be derived from deontic and/or epistemic meanings. Another option would be a development straight out of dynamic meanings (root possibility). The status of concessive meanings is unclear. Van der Auwera et al. (2009) note that concessive meanings may be a subtype of ‘general’ epistemic possibility or a further development. In the present study, concessive meaning is classified as a postmodal meaning, i.e. a further development. Epistemic meanings may lead to concessive constructions like *She may jog, but she sure looks unhealthy to me* (example from Traugott & Dasher 2002:115). In these clause types of the general form ‘although p, q’, the inferred conclusion (‘although she may jog, as you say...’, paraphrase by Traugott & Dasher ibid.) contrasts with the speaker’s own opinion. Epistemic and/or root possibility may give rise to concessive constructions in which various latent possibilities are contrasted, for example ‘whatever X may be, Y’, ‘whether or not X, Y’ or ‘be it X or Y, Z.’

This latter type of concessive constructions is closely related to eventuality meanings, the status of which is unclear as well. There are two types of eventuality meanings. The first type

denotes latent possibilities. The second type resembles the postmodal meaning ‘conditional’ as mentioned in Van der Auwera & Plungian (1998:93), who illustrate this type with Dutch *mocht* ‘in case that, if’, as shown in example (93a). This second type is accompanied by the conditional marker *om* ‘in case that, if’ in Danish and Norwegian, as shown in (93b).

- (93) a. *Mocht*      *ik*      *ziek*      *worden*,      Dutch  
 Might      I      sick      become,  
*zoek dan een vervanger.*  
 search then a      substitute  
 ‘Should I get sick, look then for a substitute.’
- b. *Om*      *du*      *mätte*      *ønske*      *det.*      Norwegian  
 If      you      must<sub>PAST</sub>      wish      it.  
 ‘In case you would want it.’

For these kinds of hypothetical meanings one can think of several developmental paths. The expression of latent possibilities by *mätte* may be a subtype or specialization of general possibility, a meaning which now has been taken over by *kunne* ‘can’ or a remnant of a former *optativus-potentialis* form (Björkstam 1919:154,166). Björkstam describes two types of optative, wish and assumption. The eventuality meaning of *mätte* is, according to Björkstam (ibid), a development from an optative form expressing assumption to an auxiliary denoting potentiality.

In this section I have presented possible developmental paths for modal and postmodal meanings in MUST/MAY. Already from the earliest stages on, a wide variety of co-existing modal meanings was available that continued to develop more polysemies in the domain of necessity, possibility and beyond. Since this study is primary synchronic in nature, I can only sketch plausible scenarios for the semantic development of MUST/MAY in broad outlines. A detailed diachronic study may reveal the micro steps that are involved.

### 3.4 Sources and method

Random samples of sentences with *må*, *mätte* and *måste* have been taken from the modern online corpora *KorpusDK*, *Norsk Aviskorpus* and *Språkbankens konkordanser* (see Section 1.4.2). For Swedish, both *må*, *mätte* and *måste* will be compared to Danish and Norwegian *mätte* because they are all related, but separate modals. For Danish and Norwegian both the present tense, *må*, and the past tense, *mätte*, are investigated as the past tense may be used to convey modal dimensions (e.g. Perkins 1983). For Swedish, samples of the forms *må*, *mätte* and *måste* have been taken because these modals have defective inflectional paradigms and do not clearly distinguish between present and past tense, see Section 3.3 and Table 11. for more details.

In *KorpusDK* there are 71,814 hits for *må* (present tense) and 25,810 hits for *mätte* (past tense). Random samples of 500 sentences have been taken for both *må* and *mätte*. *Norsk Aviskorpus* contains 814,457 hits for *må* (present tense) and 253,166 hits for *mätte* (past tense). Random samples of 500 sentences have been taken for both *må* and *mätte*. In *Språkbankens*

*konkordanser* there are 4,782 hits for *må*, only 735 hits for *mätte* and as many as 133,478 hits for *måste* in the entire corpus. These numbers clearly show that *må* and *mätte* became restricted to specific contexts after they were replaced by *måste*. Random samples of 1000, 134 and 500<sup>54</sup> sentences with *må*, *mätte* and *måste* respectively, have been taken from the subcorpora P95-98.

Only declarative clauses in which *må*, *mätte* and *måste* function as auxiliaries are considered. Infinitival forms, perfect forms, questions, modal combinations (e.g. Norwegian *vil måtte* ‘will must’) are excluded from the analysis, as in these cases it is not possible to determine the meaning of *må*, *mätte* and *måste* independent of other factors. These clauses are dismissed and form the category ‘excluded’. The corpus data are classified according to their different meanings and uses, i.e., deontic, epistemic, optative, concessive or eventuality meanings, as illustrated in example (90). Classifying modal meanings is, however, by no means a straightforward task. Especially Swedish *må* is an elusive linguistic item, whose exact meanings can be hard to depict because of subtle overlapping nuances at the interface of necessity and possibility. Context and clause-internal clues (e.g. modal particles, adjunct and modifying elements) are used to find out which reading is most plausible for ambiguous instances of *må*, *mätte*, and *måste*.

The classified corpus data, i.e., the counts per semantic category, are analyzed by means of a Chi-square test. This is a statistical method to check whether or not there is an association between two categorical variables (Field 2005:682-702). That is, in this case, if there is a statistical relation between language and the semantic distribution of a modal. If there is a relation between language and the semantic distribution of a modal, the semantic distributions of *må*, *mätte* and *måste* are significantly different for Danish, Norwegian, and Swedish. If there is no relation between language and the semantic distribution of a modal, the semantic distributions of *må*, *mätte* and *måste* are not significantly different in the Mainland Scandinavian languages. Thus, if Danish, Norwegian and Swedish display similar degrees of grammaticalization for *må*, *mätte* and *måste*, these modals express the same types of meaning in similar proportions. If the Mainland Scandinavian languages represent different degrees of grammaticalization for these modals, *må*, *mätte* and *måste* express different types of meanings in dissimilar proportions.

Values for Cramer’s V, a measurement to determine the strength of an association between categorical variables (Field 2005: 689, 693), are also given since Chi-square only states whether or not there is a significant association between two categorical variables. It does not provide information about the strength of a statistical relationship. Cramer’s V is always between 0 and 1; 0 means that there is no association at all, 1 means that there is a perfect association, and values larger than 0.5 indicate a strong relation between two categorical variables.

---

<sup>54</sup> This study started out with a sample of 1000 occurrences of Swedish *må*. It turned out that Swedish *mätte* occurs very infrequently, therefore all occurrences in the subcorpora (=134) are part of the sample. For Danish and Norwegian *mätte* a total of 1000 occurrences is analyzed, 500 for the present tense and 500 for the past tense. Since Swedish *måste* does not clearly distinguish between the present and the past tense, a sample of 500 occurrences has been taken.

### 3.5 Results

This section presents the results of the comparative corpus investigation of the modals *må*, *måtte* and *måste*. Section 3.5.1 discusses the results of the comparison for Danish *må* (present tense), Norwegian *må* (present tense), Swedish *må* (originally present tense) and Swedish *måste* (unclear tense). Section 3.5.2 reports on the results of the comparison for Danish *måtte* (past tense), Norwegian *måtte* (past tense), Swedish *måtte* (originally past tense) and Swedish *måste* (present or past tense).

Note that the sample of *måste* cannot be directly compared to Danish and Norwegian *måtte* because it is indeterminate with regard to tense inflections, while Danish and Norwegian *måtte* do inflect for tense. Since Swedish *måste* can be considered to be the present-day equivalent of Danish and Norwegian *måtte*, its semantic distribution will be presented to give an overview of the possible meanings of *måste*, regardless of its tense.

#### 3.5.1 Results for *må* (present tense)

The semantic distributions of Danish *må*, Norwegian *må* and Swedish *må* and *måste* are shown in Figure 5 below. See Appendix 1, for an overview of the counts per semantic category in each language.

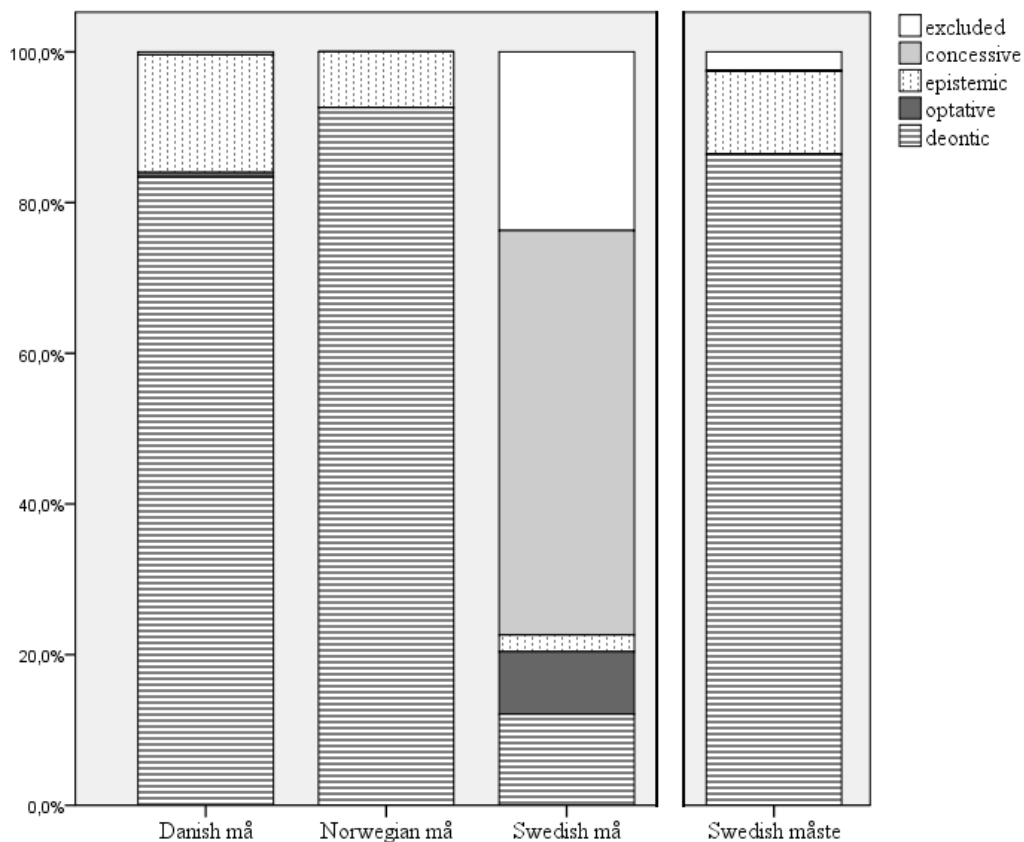
The semantic distribution of *må*<sup>55</sup> is significantly different in Danish, Norwegian and Swedish ( $\chi^2=1510$ ,  $df=8$ ,  $p<0.01$  and Cramer's  $V=0.614$ ), which indicates different degrees of grammaticalization in the Mainland Scandinavian languages. The high value of Cramer's  $V$  signifies a strong association between type of meaning and language. It points to a language-specific development for Swedish *må*, the distribution of which is clearly different from Danish and Norwegian *må*.

The semantic distributions of Danish and Norwegian *må* and their Swedish equivalent *måste* is also significantly different ( $\chi^2=17.13$ ,  $df=2$ ,  $p<0.01$  and Cramer's  $V=0.108$ ).<sup>56</sup> The low value of Cramer's  $V$  indicates a weak association for language and meaning pointing to a more general development in all languages.

---

<sup>55</sup> For Swedish *må* lexical meaning constitutes 22.9% of the sample (=excluded).

<sup>56</sup> To meet the requirements of a Chi-square test, categories that contained no or too few counts have been left out in order to carry out a reliable analysis. This test is therefore only based on the distribution of deontic and epistemic meanings.



**Figure 5.** Bar chart of the semantic distribution of *må* (present tense) and *måste*.

The different meanings of *må* and *måste* will each be discussed by means of corpus examples in 3.5.1.1-4 below. The overall picture that arises from the semantic distributions is summarized in 3.5.1.5.

### 3.5.1.1 Deontic meanings

Deontic meanings (Section 1.1.2) are by far the most frequent meanings of Danish and Norwegian *må* and Swedish *måste*. They constitute 83.4% of the Danish sample, 92.6% of the Norwegian sample, and 12.1 % and 86.4 % of the Swedish samples involving *må* and *måste*, respectively. These high frequencies were to be expected as notions like necessity, obligation, permission and (moral) desirability are prototypical for modals meaning ‘must, may.’

The permission meaning is frequent in Danish, as *måtte* can have either permission or an obligation meaning (cf. (89)). In Swedish the permission meaning occurs in archaic constructions and in Norwegian in fixed expressions. The deontic meanings of Swedish *må* are predominantly

older, archaic uses, i.e., permission meanings or more or less fixed expressions. Corpus examples of deontic meanings of *må* and *måste* are given in (94).

- (94) a. *Vi må se realiteterne i øjnene.* Danish  
 We must see realities.the in eyes.the.  
 ‘We have to face reality.’
- b. *Jeg må innrømme at vi var heldige.* Norwegian  
 I must admit that we were lucky.  
 ‘I must admit that we were lucky.’
- c. *Men man må hålla i minnet att 1995 generellt* Swedish  
 But one must hold in mind.the that 1995 generally  
*var ett synnerligen sorgesamt år för hela detaljhandeln.*  
 was an extremely miserable year for entire retail trade.the  
 ‘But one has to keep in mind that 1995 in general was a very miserable year for the entire retail trade.’
- d. *Vi måste utreda mera säger polisen.* Swedish  
 We must investigate more says police.the.  
 ‘We have to investigate more says the police.’

### 3.5.1.2 Epistemic meanings

Epistemic meanings (Section 1.1.3) are concerned with speaker judgments on degrees of possibility and probability. The modals *må*, *måtte* and *måste* express epistemic necessity which encompasses high degrees of probability, logical conclusion or inference. Epistemic meanings occur frequently in constructions like ‘must be X’ or ‘must have been X’ and are often accompanied by modal particles that reinforce the degree of uncertainty of the utterance even more. Examples of these co-occurring particles are *da* ‘thus, then’ or *vel* ‘well’, as shown in the Danish example (95a). Epistemic meanings are present in all samples and constitute 15.6% of the Danish sample, 7.4% of the Norwegian sample and 2.2% and 11% of the samples for Swedish *må* and *måste* respectively. Corpus examples of epistemic meanings of *må* and *måste* are given in (95).

- (95) a. *Det må da vel være logikken i det,* Danish  
 It must thus well be logic.the in it,  
*ikke sandt?*  
 not true?  
 ‘That must be the logic in it, right?’



- b. *Her må det ha vært kommunikasjonssvikt.* Norwegian  
 Here must it have been communication.lack.  
 ‘There must have been lack of communication here.’
- c. *Detta kan visa sig mer betydelsefullt* Swedish  
 This can show itself more consequential  
*än man i förstone må ha varit*  
 than one at first might have been  
*benägen att tro.*  
 inclined to think.  
 ‘This may be of greater importance than one at first  
 might have been inclined to think.’
- d. *Det måste finnas något problem där.* Swedish  
 It must be some problem there.  
 ‘There must be some problem there.’

### 3.5.1.3 Optative meanings

Optative meanings are about wishes and desires of the speaker and are only part of the Danish and Swedish samples. Norwegian *må* may express optative meaning (Faarlund et al. 1997) but this use seems to be too infrequent to be part of the sample. Swedish *måste* cannot express optative meaning (Teleman et al. 1999).

In optative expressions, *må* often occurs as the first constituent of the clause but may also occur in subordinate clauses or co-occur with verbs expressing hope or desire. Optative meanings constitute 0.6% of the Danish sample and 8.3% of the Swedish sample. Corpus examples of optative meanings of *må* are given in (96).

- (96) a. *Men jeg håber at denne dag [...] også må* Danish  
 But I hope that this day also may  
*bringe held og lykke til mit lidende folk.*  
 bring luck and happiness to my suffering people.  
 ‘But I hope that this day also will bring luck and happiness  
 to my suffering people.’
- b. *Må det inte regna på festdeltagarna.* Swedish  
 May it not rain on party-goers.  
 ‘Hopefully it won’t rain on the party-goers.’

### 3.5.1.4 Concessive meanings

Concessive constructions involve clauses that ‘indicate that the situation in the matrix clause is contrary to expectation in the light of what is said in the concessive clause’ (Quirk et al. 1985 :1089). They are of the general form in (97), of which there are various subtypes (see Crevels 2000:1).

(97) although p, q

Concessive meanings are only part of the sample of Swedish *må* and constitute by far its most frequent meaning, 53.9% of the sample. There are various specific concessive constructions and expressions in which *må* occurs. The most prototypical constructions are (98a,b), with or without the adversative connector *men* ‘but’. In addition to these constructions, there are a number of concessive expressions in which *må* occurs, like *Må det*, *Må så vara* or, *Det må vara hänt*, *Må vara att* ..., all meaning ‘so be it / be that as it may.’ Corpus examples of concessive meanings of Swedish *må* are given in (98).

- (98) a. *Man må lära av sina misstag, men*  
 One may learn of one’s mistakes, but  
*det finns alltid nya misstag att begå.*  
 there are always new mistakes to make.  
 ‘One may learn of one’s mistakes, but there are always new mistakes to be made.’
- b. *Och Madonna må heta Madonna –*  
 And Madonna may be named Madonna –  
*ett helgon blir hon aldrig.*  
 a saint becomes she never  
 ‘And Madonna may be called Madonna – she will never be a saint.’
- c. *Vad gäller frågan om/när det kommer*  
 What concerns question.the if/when it comes  
*en ny istid så är svaret faktiskt*  
 a new ice age so is answer.the just  
*att ingen egentligen vet, hur obekvämt*  
 that nobody really knows, how uncomfortable  
*detta svar än må vara.*  
 this answer -ever may be.  
 ‘Concerning the question ‘if/when there will be a new ice age’, the answer is basically that no one really knows, how(ever) unsatisfactory this answer may be.’

- d. *Luften må vara hur varm som helst under dagen,*  
 Air.the may be how warm as ever under day.the,  
*är vattnet kallt blir kvällen kall.*  
 is water.the cold becomes evening.the cold.  
 ‘However warm the air may be during the day, if the water is cold then the evening is cold.’
- e. *I Amerika är arkitektritade privatbostäder*  
 In America are architect designed private.homes  
*- det må vara flerbostadshus eller lyxiga strandvillor-*  
*- it may be multifamily.homes or luxury beach residences-*  
*undantag, inte regel.*  
 exceptions not rule.  
 ‘In America, private homes designed by architects -be it multifamily homes or luxury beach residences- are exceptions, not the rule.’

### 3.5.1.5 Overall results

Danish and Norwegian *må* and Swedish *måste* all have predominantly deontic meanings and epistemic meanings come second. This pattern is attested in Danish, Norwegian and Swedish, and seems to be most established in Danish as epistemic meaning is most frequent for Danish *må* compared to Norwegian *må* and Swedish *måste*. Postmodal meanings, i.e. optative and concessive, are most frequent for Swedish *må*, but are not common at all for Danish and Norwegian *må* which have predominantly deontic meanings.

### 3.5.2 Results for *måtte* (past tense)

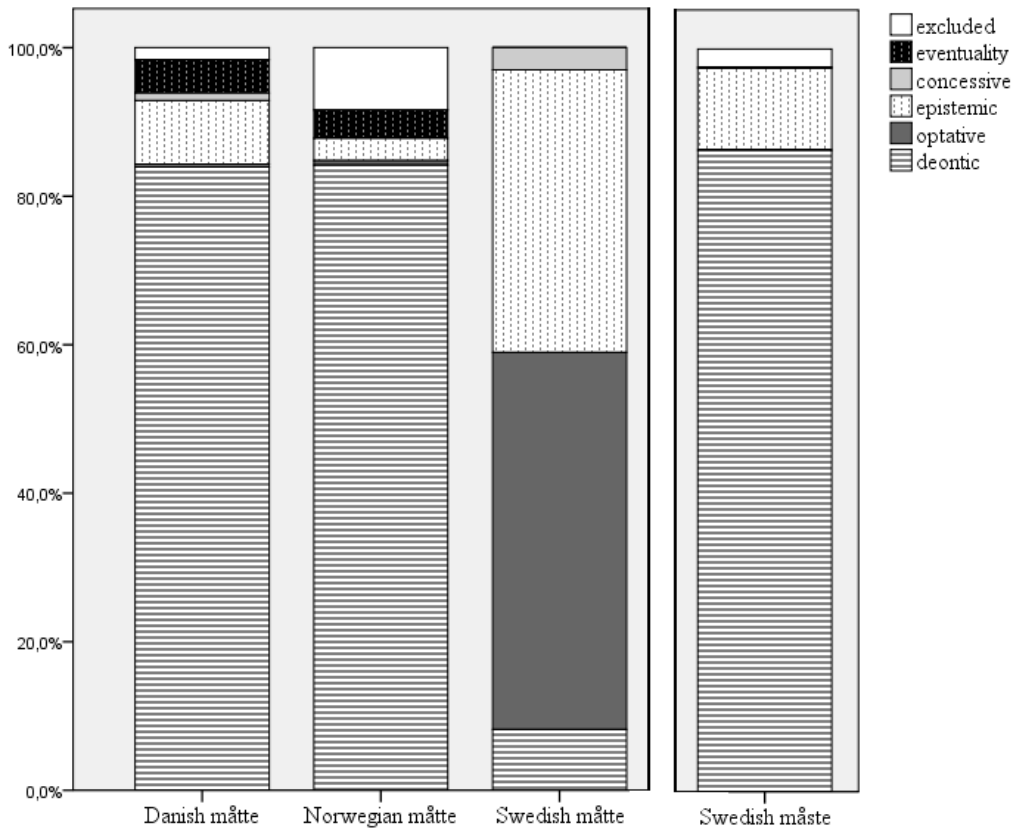
The distribution of the different meanings and uses of Swedish *måste* and Danish, Norwegian and Swedish *måtte* is shown in Figure 6. See Appendix 2 for an overview of the counts per semantic category in each language. Recall that the percentages for the semantic distribution of *måste* presented in the previous section are also compared to the results of *måtte* in this section.

The semantic distribution of *måtte* is significantly different in Danish, Norwegian and Swedish ( $\chi^2=715.4$ ,  $df=8$ ,  $p<0.01$  and Cramer’s  $V=0.564$ ), which signifies different degrees of grammaticalization<sup>57</sup> The high value of Cramer’s  $V$  indicates a strong association between meaning and language. It points to a language-specific development for Swedish *måtte*, the distribution of which is clearly different from Danish and Norwegian *måtte*.

The distributions of Danish and Norwegian *måtte* and their Swedish equivalent *måste* is also significantly different ( $\chi^2=81.82$ ,  $df=8$ ,  $p<0.01$  and Cramer’s  $V=0.166$ ). The low value of Cramer’s  $V$  indicates a weak association for language and meaning, which points to a more

<sup>57</sup> To meet the requirements of a Chi-square test, the categories optative and concessive were conflated into a single category ‘postmodal meanings’ because the separate categories contained either no counts or too few counts for a reliable Chi-square test.

general pattern for all languages. That is, the coexistence of deontic and epistemic meaning is a well-known tendency for modals meaning ‘must, may’ (Section 3.3).



**Figure 6.** Barchart of the semantic distribution of *mätte* (past tense) and *måste*.

The different meanings of *mätte* and *måste* will each be discussed by means of corpus examples in 3.5.2.1-5 below. The overall picture that arises from the semantic distributions is summarized in 3.5.2.6.

### 3.5.2.1 Deontic meanings

Deontic meanings are part of the semantic distribution of all Mainland Scandinavian languages and constitute 84.2% of the Danish sample, 84.2% of the Norwegian sample and 8.2% and 86.4% of the Swedish samples of *mätte* and *måste*, respectively. This goes to show that these modals to a large extent have retained their prototypical deontic meanings. The percentages are comparable to the results of *må*. This shows that Danish and Norwegian *må* and *mätte*

predominantly express present and past tense of deontic meanings, unlike Swedish *må* and *måtte*, which have developed into two different modals (compare Figure 5 and Figure 6). Corpus examples of deontic meanings of *måtte* are given in (99).

- (99) a. *Jeg måtte gentage det, før han reagerede.* Danish  
 I must-PAST repeat it, before he reacted.  
 ‘I had to repeat it, before he responded.’
- b. *Dagen etter måtte jeg roe litt ned.* Norwegian  
 Day.the after must-PAST I rest little down.  
 ‘The day after I had to take it easy’
- c. *Deras krav på att Arafat måtte ombilda* Swedish  
 Their claims on that Arafat must-PAST reorganize  
*sin regering har dock lämnats ohörda.*  
 his government has nevertheless left unheard.  
 ‘Their claim that Arafat must reorganize his government was nevertheless left unheard.’

### 3.5.2.2 Epistemic meanings

Epistemic meanings are present in all samples. They occur mainly in constructions like ‘must-PAST be X’ or ‘must-PAST have been X’ and are often accompanied by modal particles. Epistemic meanings constitute 8.4% of the Danish sample, 3% of the Norwegian sample, and 38.1% and 11% of the Swedish samples involving *måtte* and *måste*, respectively. Corpus examples of epistemic meanings of *måtte* are given in (100).

- (100) a. *Der måtte være forfærdelig koldt heroppe* Danish  
 It must-PAST be terribly cold here.up  
*om vinteren.*  
 in winter.the.  
 ‘It must be terribly cold up here in winter.’
- b. *Da måtte det ha skjedd noe* Norwegian  
 Then must-PAST it have happened something  
*som berørte meg personlig, sier hun.*  
 that touched me personally, says she.  
 ‘Then something must have happened that affected me personally, she says.’
- c. *Hon måtte sannerligen ha varit* Swedish  
 She must-PAST truly have been

*en       märkvärdig       dam.*  
*an       extraordinary       lady.*  
 ‘She really must have been an extraordinary lady.’

### 3.5.2.3 Optative meanings

Optative meanings are part of all samples but occur very infrequently in Danish and Norwegian, only 0.4% and 0.6% of the samples, respectively. For Swedish *måtte*, optative is by far the most common meaning, it constitutes 49.3% of the sample. In optative contexts, *måtte* is predominantly the first constituent of the clause. Corpus examples of optative meanings of *måtte* are given in (101).

- (101) a. *Måtte       det       komme       snart!* Danish  
           Must-PAST    it       come       soon!  
           ‘May it come soon!’
- b. *Måtte       han       ta       solid       og       ettertrykkelig       feil.* Norwegian  
           Must-PAST    he       take       solid       and       emphatically       wrong.  
           ‘May he be substantially and emphatically wrong.’
- c. *Måtte       hon       bli       hundra       år!* Swedish  
           Must-PAST    she       become       hundred       years!  
           ‘May she live a hundred years!’

### 3.5.2.4 Concessive meanings

Concessive meanings occur very infrequently in the Danish and Swedish samples, only 1% and 3%, respectively.<sup>58</sup> Note that the concessive meanings in the sample occur in specific constructions. For Norwegian no examples are attested in the sample, but the construction in (102) below is possible in Norwegian as well.<sup>59</sup>

A remarkable difference between Danish and Norwegian on the one hand and Swedish on the other hand is that Norwegian and Danish use past tense forms in concessive constructions, whereas Swedish uses the original present tense form *må* in concessive contexts, see also Section 3.4.1.4 above. Corpus examples of concessive meanings of *måtte* are given in (102).

- (102) a. *Hvad       beslutningen       end       måtte       blive,       bliver       det* Danish  
           What       decision.the       ever       must-PAST       be,       is       it

<sup>58</sup> This is presumably a fairly universal tendency as concessives are pretty rare and involve complex sentences.

<sup>59</sup> Separate searches in *Norsk Aviskorpus* provide examples of concessive constructions in Norwegian, the occurrence of which seems to be too infrequent to be part of the sample. An example is *Det er en historie som bygger på fakta og fakta alene, hvor irrelevante de enn måtte være.* ‘It is a story that consists of facts and facts only, however irrelevant these may be.’

*nødvendigt med en moderat og disciplineret finanspolitik.*  
 necessary with a moderate and disciplined financial policy.  
 ‘Whatever the decision may be, a moderate and disciplined financial policy is necessary.’

- b. *Framför allt går det knappast längre att hålla* Swedish  
 Above all goes it hardly longer to hold  
*tyskar, amerikaner, ryssar, fransmän, italienare och*  
 Germans, Americans, Russians, Frenchmen, Italians and  
*australiensare borta från den fornnordiska forskningen,*  
 Australians away from the Old-Nordic research.the,  
*vad än vikingakongressens brittiska veteraner*  
 what- ever viking congress’ British veterans  
*måtte* *önska.*  
 must<sub>-PAST</sub> wish.

‘Above all things, it is hardly maintainable to keep Germans, Americans, Russians, Frenchmen, Italians and Australians from Old Nordic research, whatever the British veterans of the Viking congress may wish.’

### 3.5.2.5 Eventuality meanings

Eventuality meanings are about latent or hypothetical possibilities, i.e. things that could possibly happen. Not all linguists would agree on a separate category of eventuality as this meaning could be argued to be subsumed under epistemic meanings because it concerns possibilities. However, I think this potential or hypothetical use is different from inferred conclusions as expressed by epistemic necessity. That is, eventuality does not express an evaluation of possibilities/probabilities on the basis of speaker’s knowledge.

The eventuality meaning is mentioned as a separate use of *måtte* in the literature, grammars and dictionaries (e.g. Falk & Torp 1903; Faarlund et al. 1997; Eide 2005:5). In Old Norse, potentiality used to be a third meaning besides permission (retained in Danish) and necessity (Falk & Torp 1903), the latter of which is nowadays the most frequent meaning of Danish and Norwegian *måtte*.

The eventuality meaning is only attested in the Danish and Norwegian samples, 4.4% and 3.8% respectively. In these contexts *måtte* is often accompanied by adverbs meaning ‘could possibly’, e.g. Danish *eventuel* and Norwegian *eventuell*. Corpus examples of eventuality meanings of *måtte* are given in (103).

- (103) a. *Reklamationsretten er ikke en garanti mod* Danish  
 Reclamation right.the is not a guarantee against  
*enhver fejl som måtte opstå efter købet ...*  
 every mistake that must.<sub>PAST</sub> arise after purchase.the...  
 ‘The guarantee does not cover every defect that could (possibly)  
 arise after the purchase ...’
- b. *I kveld hadde studentprest i Trondheim* Norwegian  
 This evening had student priest in Trondheim  
*Birte A. Gresseth, arrangert åpen kirke for*  
 Birte A. Gresseth, arranged open church for  
*de som måtte ønske det.*  
 those who must.<sub>PAST</sub> wish that.  
 ‘This evening, the university chaplain in Trondheim, Birte. A.  
 Gresseth, arranged open church for those who might be interested.’

### 3.5.2.6 Overall results

Deontic meanings are most common for Danish and Norwegian *måtte*, as well as Swedish *måste*. Swedish *måtte* has primarily epistemic and optative meanings. These meanings are also attested for Danish and Norwegian *måtte*, but are not that frequent compared to Swedish *måtte*. A striking difference is that eventuality is expressed by Danish and Norwegian *måtte*, but not by Swedish *måtte* and *måste*.

## 3.6 Grammaticalization and (inter)subjectification of MUST/MAY

The development of modals is generally conceived of as a prototypical instance of grammaticalization (Section 2.1.1), or more specifically auxiliatation. Modals are linguistic items on their way from lexical to grammatical status, and from grammatical to (more) grammatical status, both formally and semantically. Moreover, auxiliatation is often accompanied by subjectification. In what follows, I will discuss and analyze the development of Mainland Scandinavian MUST/MAY in relation to the definition of grammaticalization in (45) and subjectification (79) as proposed in Chapter 2, and show that *må*, *måtte* and *måste* represent different degrees of grammaticalization and (inter)subjectification.

### 3.6.1 Grammaticalization of MUST/MAY

As defined in the present study, grammaticalization is a composite type of language change whereby lexical or already grammaticalized items, in certain linguistic contexts, undergo both semantic reinterpretation and formal reanalysis. It is accompanied by a subset of correlated primitive changes and side effects that may, but need not be, involved in grammaticalization. Grammaticalization leads to a grammatical item, i.e. a linguistic item belonging to a minor category, with relational meaning, secondary status, and which regulates and organizes



grammatical structure and relations.

These different components of the definition will be separately discussed in relation to the development of Mainland Scandinavian MUST/MAY. First the essential mechanisms will be discussed in subsection 3.5.1.1, then the accompanying primitive changes will be examined in subsection 3.5.1.2, the side effects that may identify potential instances of grammaticalization are elaborated on in subsection 3.5.1.3, and finally, in 3.5.1.4 I will comment on the grammatical status and degrees of grammaticalization for Mainland Scandinavian MUST/MAY.

### 3.6.1.1 Mechanisms in the grammaticalization of MUST/MAY

In essence, the different types of language change are the result of formal reanalysis and semantic reinterpretation (Section 2.1). The development of modal auxiliaries is generally considered to be an instance of grammaticalization. This is confirmed by the shaded cells in Table 12 which show the mechanisms involved in the development of MUST/MAY.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	±
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-
<b>reinterpretation</b>					
<b>-metaphor/metonymy</b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

**Table 12.** Mechanisms in the grammaticalization of MUST/MAY.

The shift from main verb (Old Danish *mughu/muga*, Old Norse *mega*, Old Swedish *magha*) to auxiliary (Danish *måtte*, Norwegian *måtte*, and Swedish *må* and *måtte*) is a case of *categorical reanalysis* from a major to a minor category. Subsequent changes, e.g. from epistemic modal to concessive marker, are instances of intra-categorical shifts, i.e. categorical reanalysis from grammatical to (more) grammatical status. Mainland Scandinavian *må*, *måtte* and *måste* are moving towards auxiliary status. Swedish *må* and *måtte* have progressed most when it comes to

auxiliation, whereas Danish and Norwegian *måtte* still have many properties in common with main verbs. They take a middle position in between main verb and auxiliary.

The transition from deontic to epistemic meaning has often been analyzed in terms of metaphorical mapping (Bybee & Pagliuca 1985; Sweetser 1982, Heine et al. 1991). The idea is that the epistemic sense is a metaphorical extension of obligation (X is obliged to Y) to apply to the truth of a proposition (X is obliged to be true). At the micro-level semantic change is best described in terms of *metonymization*, i.e., the small, gradual steps that lead to a single change. As we have seen in Section 3.3, there are contiguous relations between premodal, modal and postmodal meanings. That is, premodal meanings give rise to modal meanings which in turn may give rise to postmodal meanings. This is witnessed by the wide variety of closely related, overlapping meanings in the domain of necessity and possibility.

The development of Mainland Scandinavian MUST/MAY captures both ‘primary’ and ‘secondary’ grammaticalization. The transformation from lexical to grammatical category and the shift from referential to relational meaning are typical of primary grammaticalization. All subsequent changes, to epistemic modal, optative or concessive marker can be subsumed under secondary grammaticalization, i.e. continued or advanced grammaticalization.

### 3.6.1.2 Primitive changes in the grammaticalization of MUST/MAY

Formal reanalysis and semantic reinterpretation are accompanied by a subset of correlated primitive changes. The accompanying primitive changes in the development of MUST/MAY are given in Table 13.

ii.	<u>Primitive changes in grammaticalization</u>	Gzn1	Gzn2
	<i>-phonology/phonetics</i>		
-	loss of phonological/phonetic substance	(+)	(+)
	<i>-morphology</i>		
-	loss of morphological compositionality	(+)	(+)
-	loss of morphosyntactic properties	+	(+)
	<i>-syntax</i>		
-	loss of syntactic variability	(+)	(+)
-	loss of syntactic autonomy	(+)	(+)
	<i>-semantics</i>		
-	loss of semantic substance	+	+
-	loss of semantic compositionality	(+)	(+)
	<i>-discourse/pragmatics</i>		
-	subjectification	(+)	(+)
-	intersubjectification	(+)	(+)

**Table 13.** Primitive changes in the grammaticalization of MUST/MAY.

As regards the level of *phonology/phonetics*, there is only loss and change of phonetic substance for MUST/MAY with respect to the creation of new infinitival and past tense forms in the preterite-present verb paradigm (Section 3.3.1). Loss of phonological/phonetic substance is closely interrelated with the domain of *morphology*, especially loss of morphosyntactic properties. In the case of Mainland Scandinavian MUST/MAY we see that the Swedish modals have the most defective inflectional paradigms, see Table 11. They lack infinitival forms and they do not inflect for tense. Swedish *måste* has a perfect form, but Swedish *må* and *måtte* lack perfect forms. Danish and Norwegian *måtte* do have infinitival forms and inflect for tense. All modals have deficient verbal paradigms in the sense that they lack present participles, imperative and passive forms. The loss of morphosyntactic properties goes hand in hand with the categorical reanalysis from main verb to auxiliary, whereby an item gradually loses the prototypical features of the major category such as inflection. There is no change with respect to the morphological compositionality of MUST/MAY because they are monomorphemic forms, not compositional ones. As such, changes that affect the morphological compositionality of a form, i.e. fusion and coalescence, do not apply.

The reanalysis from main verb to auxiliary inevitably leads to a decrease in syntactic variability. Auxiliaries are less flexible than main verbs and acquire fixed syntactic slots, i.e. they become more tightly integrated in the syntactic structure as they become dependent on the main verb of the clause. In deontic contexts the modals are free to combine with other verbs, in epistemic contexts the construction gets more restricted as the prototypical form of epistemic expressions is a stative predicate of the form ‘must be X/must have been X’. For optative contexts the syntactic position is generally restricted to the first position in the clause or subordinate clauses introduced by a complementizer. As far as concessive contexts are concerned, there are a number of specific constructions in which *må* and *måtte* convey concessive meaning. For Danish and Norwegian, the past tense is preferred to express concessive meaning and also used to express hypothetical dimensions like eventuality.

With respect to semantic changes in MUST/MAY, it can be observed that there is loss of referential meaning, but there is gain of relational and meta-linguistic meaning on the way from lexical verb meaning ‘have the power/strength’ to auxiliary form with ability, permission, possibility and necessity meanings. The meaning of these modals becomes increasingly dependent on the context in which they occur. Through the course of time the metonymic relations between the different possible meanings may fade so that an item becomes opaque. This applies especially to Swedish *må* whose exact meanings can be hard to establish. All the contiguous semantic changes have led to the extremely polysemous and opaque linguistic item that it is today.

As regards the discourse/pragmatics level, there is an increase in (inter)subjective meaning for all modals, most notably for Swedish *må* and *måtte*. Dynamic and deontic modality, also termed ‘root modality’, are traditionally distinguished from epistemic modality because they do not express speaker judgments, i.e. they do not show signs of (*inter*)*subjectification*. However, deontic modality comprises permission and obligation meanings, which do have

(inter)subjective dimensions because these are directive uses. Epistemic meanings are concerned with the speaker's subjective evaluation of the likelihood of a state of affairs. The speaker judges from his/her own perspective that something is probable or must be the case. Optative meanings concern the speaker's personal or collective wishes, which are clearly (inter)subjective.

Concessive meanings are (inter)subjective because the speaker evaluates his/her statement in light of contrasting opinions or general accepted truths. As such it may involve interaction between speaker/writer and addressee/reader in case the speaker's opinion contrasts with the interlocutor's statement in the matrix clause (e.g. although p, as you say, I think q). In case of epistemic and postmodal meanings there has been a reanalysis from syntactic to speaking subject (cf. Benveniste 1966). The role of (inter)subjectification in the development of Mainland Scandinavian MUST/MAY will be discussed in more detail in Section 3.6.2.

### 3.6.1.3 Side effects in the grammaticalization of MUST/MAY

The side effects of formal reanalysis and semantic reinterpretation and their accompanying primitive changes can be used as a diagnostic to identify potential cases of grammaticalization, as these are observable signs of ongoing change. The side effects of grammaticalization in the development of MUST/MAY are shown in Table 14.

iii.	<u>Side effects of grammaticalization</u>	Gzn1	Gzn2
-	paradigmaticization	+	+
-	obligatorification	(+)	(+)
-	condensation	+	+
-	layering, divergence, specialization, persistence	+	+
-	productivity	+	+
-	frequency	+	+
-	typological generality	+	(+)

**Table 14.** Side effects of grammaticalization in MUST/MAY.

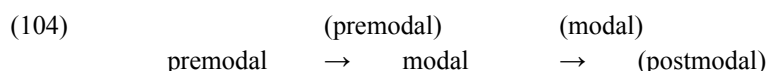
*Paradigmaticization*, i.e. the degree to which a linguistic item is part of a paradigm, corresponds to *productivity* (context expansion). That is, when a linguistic item enters another paradigm it can be used in more and different contexts, this is also known as 'host-class expansion' (Himmelfmann 2004). In case of MUST/MAY we see that they enter the paradigm of modal auxiliaries, which in turn single themselves out by being preterite-present verbs.<sup>60</sup> Once their status as a modal is established they continue to develop more modal and postmodal meanings. Consequently, when a linguistic item can be used in more contexts, i.e. is part of more paradigms, it occurs with increasing *frequency* (both type and token frequency).

*Condensation* and *obligatorification* are two parameters that are often mentioned in

<sup>60</sup> Not all preterite-present verbs are modal auxiliaries, e.g. Old Norse *eiga* 'to own' and *vita* 'to know' are exceptions.

relation to formal change. The degree to which modals are obligatory in expressing modal and postmodal meanings is hard to establish as there are alternative ways of expressing these meanings (e.g. by means of adverbs, predicative adjectives or lexical means, such as nouns). Condensation, decrease in syntactic scope and increased dependency, has been a much-debated parameter. It relates to syntactic integration and leads to a decrease in syntactic autonomy of a linguistic item. In the development of MUST/MAY its semantic scope expands (cf. the hierarchy of qualificational categories in Chapter 2) because there has been a reanalysis from syntactic to speaking subject, whereas its structural scope is reduced because MUST/MAY becomes dependent on the main verb and acquires a fixed slot in the clause.

The synchronic coexistence of more and less grammaticalized manifestations of a given form is called *layering*. Through the course of time, *må*, *mätte* and *måste* developed a wide variety of meanings within the realm of necessity and possibility. The newly emerged and older meanings (=layers) may coexist for hundreds of years, eventually some meanings may disappear, as shown in (104) below.



A subtype of layering is *divergence* or split. This means that the original word or construction from which a grammaticalized word or construction emerged, continues to exist and may undergo the same changes as ordinary lexical items. This is the case for Swedish *må* which has a full-fledged lexical variant meaning ‘to feel (bad/well)’ besides its modal variant.

*Specialization* is the phenomenon that out of a variety of near-synonym forms only one becomes prominent in expressing a certain meaning. In Old Swedish *må* and *mätte* could express all kind of modal dimensions which made them linguistic elements with fuzzy meanings. They were replaced by *måste* which came to denote necessity and obligation. The modal *kunna* took over their possibility meanings. There has also been overlap and confusion with the now obsolete auxiliary *månde* ‘may, might’ which developed into an epistemic question particle (see Chapter 4). Because of its semantic opacity, *må* and *mätte* may have lost terrain to other modals, such as *måste*, *kunna*, *månde*, and continue to live on as a highly specialized items in concessive and optative contexts.

*Persistence* pertains to the observation that traces of previous meanings can be traced back in newer forms. There are contiguous relations between the meanings ‘have the power/strength’, ability/capacity, possibility, permission, obligation, probability concession and wish. These are all closely related, overlapping meanings in the domain of necessity and possibility.

Finally, grammaticalization paths tend to be cross-linguistically replicated. As pointed out in section 3.3, the development of modal and postmodal meanings is a well-known tendency in semantic change. At least for the Germanic languages similar developments can be observed which makes that MUST/MAY conforms to the principle of *typological generality*.

#### 3.6.1.4 Grammatical status and degrees of grammaticalization for MUST/MAY

Grammaticalization leads to a grammatical item, i.e. a linguistic item belonging to a minor category, with relational meaning, secondary status, and which regulates and organizes grammatical structure and relations. This means that grammatical items, being secondary to other syntagmatically related items, cannot on their own be focused or questioned, because they are dependent on other items, both formally and semantically (cf. Boye & Harder 2012). Mainland Scandinavian MUST/MAY is formally dependent on the main verb of the clause and semantically dependent on the context and construction in which it occurs, as such it qualifies as a grammatical item.

Throughout this chapter I have regarded postmodal meanings as more grammaticalized than modal or premodal meanings. But in what ways are optative or concessive meanings more grammaticalized than their prior modal meanings?<sup>61</sup> As argued in Section 3.3, semantic change may signal increasing grammaticalization and is often accompanied by formal changes. As such, grammaticalizing items can be located at different clustering points (e.g. premodal, modal or postmodal meanings) or (e.g. main verb, auxiliary) along a cline, representing variant degrees of grammaticalization.

The formal side of determining degrees of grammaticalization for modals has been exhaustively discussed in the literature, but there are generally no strong claims on semantic change in relation to degrees of grammaticalization (e.g. Van der Auwera & Plungian 1998; Nuyts 2001). However, van der Auwera & Plungian (1998:116) note that in principle, “the degree of formal grammaticalization of postmodals could be either higher, equal or lower.” A higher degree of formal grammaticalization is what one would expect and this is indeed frequently attested. An example is optative ‘may’, whose canonical position came to be fixed to clause-initial position.

The postmodal variants of MUST/MAY are more grammaticalized because of their more complex semantics, i.e. relational and meta-linguistic meanings, which are accompanied by more restricted formal properties. That is, their syntactic and semantic autonomy is decreasing the more postmodal their meanings get. This is accompanied by the development of (inter)subjective meanings for MUST/MAY, which will be discussed in the next section.

---

<sup>61</sup> One might wonder whether a grammatical item can be more ‘grammatical’ than another grammatical item in terms of grammatical status and how one is to assess ‘more grammatical status.’ That is, whether grammatical status itself is gradable and if there are sound arguments for claims such that affixes, in principle, are more grammatical than for examples auxiliaries (cf. Boye & Harder 2012). I think that for one and the same item, within one and the same grammaticalization chain, this is possible and can be measured in term of loss in syntactic and semantic autonomy (cf. Lehmann 1995 [1982]). For unrelated developments and items, with their particular inherent features this may be a case of comparing apples with oranges. When considering the development of MUST/MAY the postmodal variants are more grammatical(ized) than for example the (pre)modal variants because they are the least autonomous items in the developmental chain.

### 3.6.1.5 (Inter)subjectification of MUST/MAY

As defined in the present study, subjectification and intersubjectification are metonymic types of semantic change that lead to increased speaker-perspective/involvement (subjectification) and increased focus on speaker-addressee interaction (intersubjectification), see Section 2.1.4. As mentioned several times before, the development of epistemic meanings always involves subjectification, i.e. “[...] the development of a grammatically identifiable<sup>62</sup> expression of speaker belief or speaker attitude to what is said” (Traugott 1995:32), but what kind of subjectification is involved in the development of MUST/MAY?

Table 15 summarizes the types of subjectification involved in the development of MUST/MAY.

<b>(Inter)Subjectification</b>	<b>MUST/MAY</b>
<b>I. subjectification</b> [speaker perspective, attitude and judgment]	+
- <b>textual level</b> [meta-linguistic meanings] epistemic, concessive, optative	
<b>III. intersubjectification</b> [interaction with interlocutor]	+
- <b>textual level</b> directive meanings (permission, obligation), concessive, optative	

**Table 15.** Types of subjectification in MUST/MAY.

The development of epistemic meanings is subsumed under Tendency III in semantic change (Section 2.1.4), i.e. the speaker’s subjective belief state/attitude toward the proposition. The development of epistemic meaning in MUST/MAY is an instance of grammaticalization accompanied by subjectification at the textual level. The rise of postmodal meanings is a case of grammaticalization and (inter)subjectification at the textual level.

### 3.7 Summary, discussion and conclusions

Significant differences, both formally and semantically, have been found with respect to the development of modal and postmodal meanings in the Mainland Scandinavian modals *må*, *måtte* and *måste*. These modals display different degrees of grammaticalization and (inter)subjectification, which is most clearly manifested in Swedish *må* and *måtte* which have reached the final stages of grammaticalization, whereas Danish and Norwegian *måtte* are still in full development (both formally and semantically).

<sup>62</sup> This definition implies that one can identify subjectification on syntactic grounds. A subsequent question that arises is how one is to distinguish different types of subjectification in one and the same item, assuming that, subjectification and intersubjectification are diachronically related (Traugott 2010). For example, concessive *while* which is diachronically related to temporal *while* ‘during’ concessive *while*, must bear hallmarks of both textual and interpersonal subjectification.

Danish *måtte* has primarily deontic meanings in both the present and past tense. Epistemic meanings seem to be more established in Danish compared to Norwegian *måtte* and Swedish *måste*. Postmodal meanings like concessive, optative and eventuality may be expressed by *måtte* but are far less frequent than deontic and epistemic meanings. The inflectional paradigm of Danish *måtte* is deficient in the sense that it lacks present participles and passive and imperative forms. Infinitival and perfect forms occur far less frequent in the sample than the corresponding forms of Norwegian *måtte*. Postmodal meanings occur for the most part in specific constructions.

Norwegian *måtte* has predominantly deontic meanings in both the present and past tense. Postmodal meanings like optative and concessive (not attested in the sample) are possible, but occur infrequently compared to deontic and epistemic meaning. The inflectional paradigm of Norwegian *måtte* lacks passive and imperative forms as well as present participles. Infinitival and perfect forms are very frequent and postmodal meanings occur mainly in specific constructions

Swedish *må* expresses mostly concessive meanings and Swedish *måtte* has predominantly epistemic and optative meanings. Both *må* and *måtte* have defective paradigms because they lack tense marking, present participles as well as infinitival, perfect, imperative and passive forms. The deontic meanings of Swedish *må* concern primarily older, archaic uses, i.e., permission meanings. Its epistemic meanings are highly context-dependent and hard to distinguish from related deontic and concessive meanings. Deontic and epistemic meanings are now expressed by *måste*, which has replaced *må* and *måtte*. The modals *må* and *måtte* continued to exist and have developed into highly grammaticalized and (inter)subjectified linguistic elements denoting (post)modal dimensions and occur for the most part in set phrases, specific constructions and contexts.

Swedish *måste* is formally a grammatical item<sup>63</sup> because of its defective inflectional paradigm, but semantically, it is a deontic modal that may also express epistemic modality. Since *måste* is a borrowed modal form, one might wonder whether it ever displayed a full inflectional paradigm. More historical research on the introduction and spread of *måste* in the Swedish language is needed in order to get a clearer picture of the evolution of its inflectional properties. The modals *må* and *måtte* have lost their inflectional properties in the course of time, but other Swedish core modals (e.g. *kunna* ‘can’ or *böra* ‘should’) are like Danish and Norwegian *måtte*. They have infinitival and perfect forms as well as tense marking, but they also lack present participles and imperative and passive forms.

Swedish *må* and *måtte* are semantically and formally more grammaticalized than Danish and Norwegian *måtte*. They have mainly (post)modal meanings, defective inflectional paradigms and occur primarily in specific constructions and expressions. Swedish *måste* is formally more grammaticalized than its Danish and Norwegian counterpart *måtte* because of its defective inflectional paradigm. Semantically Swedish *måste* and Danish and Norwegian *måtte* have the same degree of grammaticalization. They all have mainly deontic and epistemic meanings.

---

<sup>63</sup> The English modal *must* has a comparable status.



These findings conform to Van der Auwera & Plungian's (1998:115) observation that in the domain of modality semantic change seems to precede formal grammaticalization. Swedish *måste* might seem to contradict this tendency, because it is unclear if *måste* ever had a full inflectional paradigm. If *måste* did not have a full inflectional paradigm, it could not lose the verbal properties that are typical in a process of auxiliatization. This then explains why it is formally more grammaticalized than one would expect on the basis of its semantic distribution.

Traugott & Dasher (2002:115) hypothesize that 'since the metatextual meanings are the most (inter)subjective, they are probably the latest to develop semasiologically.' Especially the development of Swedish *må* and *måtte* (see Andersson 2007:191-194) is a good illustration of the observed tendencies for meaning change and (inter)subjectification in the modal domain. In short, in Old Swedish the dominating meanings of *må* and *måtte* are ability/capacity and permission. In Early Modern Swedish, many root modal meanings of *må* and *måtte* were lost, i.e., ability/capacity and permission meanings. The uses of *må* become more idiomatic and there is a decrease in frequency for *må* and *måtte*. Concessive meanings are now common in fixed constructions. The use of *må* and *måtte* seems to be bound to the perspective of the speaker and occurs mainly in dialogue. Optative meanings are used more and more in order to express feelings and emotions, for example in exclamations. Optative meanings occurred rarely in Old Swedish. In sum, over the course of its existence, the meanings of *må* and *måtte* have become increasingly (inter)subjective, expressing the speaker's personal views and/or emotions, as well as reference to speaker-writer and addressee-reader interaction in dialogue and exclamations (cf. also Björkstam 1919).

All in all, the development of modal and postmodal meanings has thus led to higher degrees of grammaticalization for the Mainland Scandinavian modals *må*, *måtte* and *måste*, which is reflected in more complex semantics, especially more (inter)subjective meanings, and accompanied by more restricted formal properties. The development of Danish *måtte*, Norwegian *måtte* and Swedish *må*, *måtte* and *måste* is in line with observed developmental tendencies for its Germanic cognates. That is, these modals have the ability to express various modal and postmodal dimensions and are moving towards auxiliary status.

There is also quite some language-specific variation. Swedish borrowed 'must' from Middle Low German and retained a lexical variant of *må*. In the Swedish samples there were no instances of eventuality meanings which are common for Danish and Norwegian *måtte*. Danish and Norwegian *måtte* resemble each other very much. This could be due to the fact that Danish has been the official language in Norway during the union between the two countries (until 1814), cf. Section 1.3.

The overall picture is that the closely related Mainland Scandinavian languages generally conform to observed cross-linguistic tendencies, but they display different degrees of grammaticalization and manifest other language-specific developments. This goes to show that language change is by no means deterministic. General tendencies can be identified, but changes do not have to occur and if a change does take place it does not have to go through all transition stages along a cline.

## Chapter 4

### 4. Case study II: Modal sentence adverbs MON and MAYBE

#### 4.1 Introduction

This chapter presents a comparative corpus investigation of the formal and semantic properties of the interrogative/epistemic adverb ‘I wonder’ and the epistemic adverb ‘maybe’ in the Mainland Scandinavian languages. These are Danish *mon*, Norwegian *mon* and Swedish *månne*, henceforth Mainland Scandinavian MON, and Norwegian *kanskje*, Swedish *kanske* and Danish *måske*, henceforth Mainland Scandinavian MAYBE. The focus of this chapter is the development of epistemic adverbs in relation to the lexicalization-grammaticalization-pragmaticalization interface and (inter)subjectification.

The aim of this chapter is threefold: i) to discuss the status of intra-categorical changes, ii) to examine the lexicalization-grammaticalization-pragmaticalization interface by means of empirical data from comparative corpus investigations of MON and MAYBE, and iii) to explain the peculiar syntactic status of MON and MAYBE. These adverbs may violate the Verb Second principle because they may occur in second position and/or impose subordinate word order on the syntactic structure of a clause. This atypical syntactic behaviour can be explained through the diachronic development of these adverbs, more specifically their verbal origin, see Section 4.3 and 4.4 for more details.

The development of Mainland Scandinavian MON involves a shift from auxiliary to interrogative adverb and successive intra-categorical shifts in the adverbial domain. Intra-categorical shifts cannot be explained within traditional grammaticalization analyses as they do not result in more grammatical(ized) linguistic items, i.e. there is no shift from lexical to grammatical status or from grammatical to more grammatical status. The interrogative adverb MON also came to function as an epistemic sentence adverb, and eventually as speech act adverb too. The question is whether an epistemic sentence adverb is more grammatical(ized) than an interrogative adverb. A concomitant problematic issue that arises is how to assess ‘more grammatical(ized) status.’ Moreover, on the one hand MON is an optional formulaic expression, on the other hand it is highly (inter)subjective in its speech-act use. In this respect, development of MON also has characteristics of lexicalization and pragmaticalization.

The status of Mainland Scandinavian MON will be examined on the basis of the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization as represented by Table 10 in Chapter 2. In addition, the role of (inter)subjectification in the development of MON will be examined. The following research questions, as stated in (105), will be addressed with respect to the interrogative/epistemic adverb MON.

- (105) a. What are the differences and similarities between Danish, Norwegian and Swedish with respect to the etymology, development, syntactic and semantic properties of the interrogative/epistemic adverb MON?

- b. Does the development of Mainland Scandinavian MON, which displays intra-categorical shifts, qualify as an instance of grammaticalization? Or, alternatively, does it qualify as an instance of lexicalization or pragmaticalization?

The development of Mainland Scandinavian MAYBE bears many hallmarks of grammaticalization, but it is not a straightforward instance of grammaticalization. Some properties of its development are typical of lexicalization too, especially changes affecting the morphological and semantic compositionality of MAYBE. Moreover, the development of unverbated epistemic adverbs has been labelled grammaticalization (Andréasson 2002; Brinton & Traugott 2005), but also characterized as lexicalization (Ramat 2001). The double-classification for this type of adverb presupposes similarities between these types of language change (cf. Section 2.2.1). In addition, MAYBE may function as a speech-act adverb. As such, adverbs of the MAYBE-type form an excellent case to study the lexicalization-grammaticalization-pragmaticalization interface.

The status of Mainland Scandinavian MAYBE will be examined on the basis of the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization as shown in Table 4 in Chapter 2. The role of (inter)subjectification in the rise of MAYBE will also be discussed. The following research questions, as stated in (106), will be addressed with respect to the epistemic adverb MAYBE.

- (106) a. What are the differences and similarities between the Mainland Scandinavian languages with respect to the etymology, development and syntactic and semantic properties of the epistemic adverb MAYBE? More specifically, do the Mainland Scandinavian languages differ in the degree of grammaticalization, lexicalization and/or pragmaticalization for MAYBE?
- b. Is the development of Mainland Scandinavian MAYBE a case of lexicalization, grammaticalization, pragmaticalization or a mixture of these?

This chapter is organized as follows. Section 4.2 discusses adverbs in general and the properties of epistemic adverbs more specifically. The comparative corpus investigations of the interrogative/epistemic adverb MON and the epistemic adverb MAYBE are presented in Section 4.3 and 4.4 respectively. Section 4.5, finally, contains a summary, discussion and the main conclusions of the case studies on modal adverbs, in which the differences and similarities in the development of MON and MAYBE will be highlighted.

## 4.2 Adverbs

The category of adverbs in general is very hard to define. Adverbs come in various types and may have different functions. In essence, an adverb is a modifier of verbs or verb phrases, which

is inherent in its name ad-verb. But adverbs may also modify adjectives, sentences, and occasionally even nouns.

There is a range from relatively open class, derived adverbs like *fortunately* to relatively closed class, monomorphemic adverbs like *now*, *just* (Ramat and Ricca 1994; 1998). Because adverbs are so diverse and internally differentiated, the category has to be regarded as a gradient one. Brinton & Traugott (2005:133) note that “it appears then that adverbs may range from contentful, lexical typologically non-prototypical adverbs that are members of an open class, to grammatical, typologically prototypical close class adverbs.” Thus, ‘adverb’ is not simply a lexical or grammatical category, rather it comprises subsets of lexical, grammatical and mixed forms (=forms that have partly concrete and partly abstract meaning). In addition, adverbs may (simultaneously) function at different layers, i.e. ideational, textual or interpersonal level. All this makes adverbs particularly suitable to explore the lexicalization-grammaticalization-pragmaticalization interface, the status of intra-categorical shifts, as well as the role of (inter)subjectification.

#### 4.2.1 Epistemic adverbs

Epistemic sentence adverbs, like sentence adverbs in general, modify the content of a clause at the sentence level. They have the entire clause as their scope. A prototypical property of sentence adverbs is their multipositionality or syntactic flexibility. That is, they may occur in various syntactic positions, i.e. they may take sentence-initial, -medial or -final positions. Epistemic adverbs form a relatively closed subclass of sentence adverbs that express various degrees of likelihood, i.e. (un)certainty, (im)possibility and (im)probability (Section 1.2).

It has been claimed that, unlike negative modal adjectives, negative modal adverbs do not exist (Bellert 1977; Nuyts 2000:107), see also the epistemic continuum in Table 1 for the degree of polarity of different epistemic expressions. For example, the adjective *probable* has the corresponding negative form *improbable*, but the adverb *probably* does not have corresponding negative form *\*improbably*. Ramat & Ricca (1998) argue against this claim by stating that negative modal adverbs do exist, but that the attested ones do not have corresponding adjectival forms.<sup>64</sup> Examples are German *keineswegs* and its corresponding forms in Dutch *geenszins*, Swedish *ingalunda*, and Danish/Norwegian *ingenlunde* ‘by no means, not at all’. Thus negative modal adverbs do exist, but they are very rare.

Another widespread claim about modal adverbs is that they cannot occur in questions (Foley & Van Valin 1984; Hengeveld 1988:236), that is, modal adverbs cannot be the first constituent in an interrogative clause or be the questioned constituent. The examples in (107) are taken from Nuyts (2001:58).

---

<sup>64</sup> Swedish *omöjligen* ‘impossibly’ is an exception to their claim, because it does have a corresponding adjectival form *omöjlig* ‘impossible.’

- (107) a.     \**Probably he is home?*
- b.     \**Is he probably home?*

Modal adverbs do occur in questions both in English and in other languages. For example, in Dutch (108a) or Norwegian (108b), ‘maybe’ regularly occurs in interrogatives and Mainland Scandinavian MON only occurs in interrogatives, as shown in the Danish and Swedish examples in (108c) and (108d) respectively.

- (108) a.     *Heeft u       **misschien**   een   paperclip   voor   me?*                   VD  
               Have you   maybe       a   paperclip   for   me?  
               ‘Do you happen to have a paperclip for me?’
- b.     ***Kanskje**       det   er   en   løsning?*                   NAK  
               Maybe       that   is   a   solution?  
               ‘Could this be a solution?’
- c.     *Hvem har   **mon**   spist   kagen?*                   PNO  
               Who has   MON eaten cake.the?  
               ‘I wonder who has eaten the cake.’
- d.     ***Måne**       detta   är   vad   hon   har   tänkt       sig?*   NSSO  
               MON       this   is   what   she   has   thought       herself?  
               ‘I wonder if this is what she expected.’

The question is whether modal adverbs in interrogative clauses express epistemic modality or not, as interrogatives inherently contain an element of uncertainty. Modal adverbs in questions are probably best analyzed as speech-act adverbs, as discussed by Nuyts in (109) below.

- (109) [...] speech act use is no doubt derived from the epistemic use [...] That is, rather than indicating a degree of likelihood of the state of affairs, the adverb modifies the ‘tendency’ of the speech act: it turns a neutral question into a tendentious one (Nuyts 2001:58).

According to Ramat & Ricca (1998) there are essentially four types of epistemic adverbs. These are *derivational adverbs* via productive word formation rules (e.g. Dutch *mogelijk-erwijs*, Norwegian *mulig-vis* ‘possible-wise’), *conversions* or *zero derivations* that may function as either an adjective or adverb (e.g. Dutch *waarschijnlijk* ‘probable/probably’), *univerbations* like for example German *vielleicht* or French *peut-être* ‘maybe/perhaps’, and borrowings as for example Saami *kanske* from Norwegian or Swedish.

These four types can be reduced to two basic types, namely derivations and

univerbations. Derivations include all forms derived from adjectives as well as zero derivation, and univerbations include all mergers of two or more formerly autonomous words into one word. Cross-linguistically the same sources and patterns have been replicated for both derived (e.g. *sann-syn-lig*, *waar-schijn-lijk* ‘true-seem-ly’) and unverbated epistemic (e.g. *may be*) adverbs.<sup>65</sup> Ramat & Ricca (1998) note that items that display morphological regularity and semantic transparency (e.g. *probably* KB) represent a less frequent alternative to non-derived or semantically opaque words (e.g. *perhaps* KB). With regard to univerbations there is a continuum from plain transparency (*maybe*) to full opacity (Dutch *misschien* < ‘*t mach schien* ‘it may happen’).

Epistemic adverbs may optionally be followed by a complementizer, for example French *probablement (que)*, Dutch *misschien (dat)*, Norwegian *kanskje (at)* or have an incorporated complementizer like Serbo-Croatian *možda* ‘can that’. Ramat & Ricca (1998:233) write that adverbs with complementizer are untypical because they do not develop from a constituent-like unit. Prosodically, they argue, in sentences like ‘it can be that’/ ‘I think that’ a pause comes after, and not before the complementizer. Therefore univerbations that retain the complementizer must necessarily derive from main predicates; those without complementizer could derive from both main clauses and parenthetical clauses. Phrases like ‘*may/can (be/happen) (that)*’ are a productive source for adverbs meaning ‘maybe, perhaps’. This pattern is attested in many Indo-European languages.

The concept of *knowing* has given rise to Spanish *quizá* ‘perhaps’ originally meaning ‘who knows’ but is also found in adverbs expressing certainty like Dutch *gewis*, Mainland Scandinavian *vi(s)st*, Swedish *förvisso* ‘certainly’. The notion of *chance* is at the basis of Latin *forsit* (> *fors sit* ‘the chance be’), *forsitan* (< *fors sit an* ‘the chance be that), Irish (*tá seans* literally ‘is chance’, and English *perhaps* (< *hap* ‘chance’, cf. *happen*, the archaic adverb *mayhap(pen)*) and the archaic adverbs *perchance*, *peradventure*. Spanish *tal vez* and Portuguese *talvez* ‘perhaps’ originally meant ‘sometimes’. Dutch *soms* ‘sometimes’ may also be used epistemically meaning ‘perhaps’, but only in questions and conditionals.

Subordinate conditional clauses may be the source to ‘perhaps’, for instance Breton *marteze*, literally ‘if-comes-this’. The adjective ‘easy’ is part of formations like German *vielleicht* (< MHG *vil Ilhte* ‘very easy/easily’) and Dutch *wellicht* ‘maybe, possibly’ and *allicht* ‘very probable, obviously’. Combinations of different source concepts is also possible, for example in Saxon dialects *lichtkaans* ‘perhaps’ literally ‘easy-chance’ or Middle Dutch *machlichte* ‘perhaps, (it) can/may easy’ which seems to be a mixture of *wellicht* and *misschien*.

#### 4.2.2 Epistemic adverbs in the Mainland Scandinavian languages

Basically, there are two types of modal adverbs that express epistemic modality in the Mainland Scandinavian languages. These are derivations from adjectives and various types of univerbations. Some examples are given in (110) below.

<sup>65</sup> All examples in this section are taken from Ramat & Ricca (1998).

- (110) a. DK *vist/sikkert* ‘certainly’, *muligvis* ‘possibly’, *sandsynligvis* ‘probably’  
 NO *visst/sikkert* ‘certainly’, *muligens* ‘possibly’, *sannsynligvis* ‘probably’  
 SE *visst/säkert* ‘certainly’, *möjligen* ‘possibly’, *sannolikt* ‘probably’
- b. DK *måske*, (*kanske*<sup>66</sup>), *kan hænde* ‘maybe/perhaps’, *monstro* ‘I wonder’  
 NO *kanskje*, *kan hende* ‘maybe/perhaps’, *monstro* ‘I wonder’  
 SE *kanske*, *kanhända*, *måhända* ‘maybe/perhaps’, *måntro* ‘I wonder, perhaps’

The present study focusses on adverbs of the latter type, i.e. Mainland Scandinavian MON (i.e. Danish *mon*, Norwegian *mon* and Swedish *månne*) and MAYBE (Norwegian *kanskje*, Swedish *kanske* and Danish *måske*). Epistemic adverbs of the MAYBE-type are by far the most frequent epistemic adverbs in the Mainland Scandinavian languages. They occur much more frequent than their derived near-synonymous forms, as shown in Table 16 below.

	Danish	KDK	Norwegian	NAK	Swedish	SK
‘maybe’	<i>måske</i>	42,068	<i>kanskje</i>	247,562	<i>kanske</i>	83,783
‘probably’	<i>sandsynlig vis</i>	2,600	<i>sannsynlig vis</i>	395	<i>sannolikt</i>	8,091
‘possibly’	<i>muligvis</i> -	2,564	<i>muligvis</i> <i>muligens</i>	103 21,525	- <i>möjligen</i>	7,389
‘I wonder’	<i>mon</i>	3,445	<i>mon</i>	1,419	<i>månne</i>	331
‘I wonder’	<i>monstro</i>	42	<i>monstro</i>	15	<i>månnro</i> <i>måntro</i>	29 12

**Table 16.** Frequencies for derived and unverbated epistemic adverbs in the Mainland Scandinavian languages.

In this section I will reflect on the peculiar syntactic status of MON and MAYBE. That is, these adverbs may violate the so-called Verb Second principle. Verb Second (henceforth V2) is a property of declarative main clauses and *wh*-questions which is found in all Germanic languages except English. In V2 languages, the finite verb is always the second constituent in a declarative main clause, i.e. only one constituent may precede the finite verb.

To illustrate the structure of V2 clauses, some examples are placed in Platzack’s scheme for main and subordinate clauses, which is based on Diderichsen’s (1946:186) classical schemes for the analysis of main and subordinate clauses in the Mainland Scandinavian languages (Platzack 1998:93). Platzack merged Diderichsen’s schemes for main and subordinate clauses into one scheme and added the ‘type’-field. The first field in this scheme is the ‘foundation’, abbreviated as ‘F’, which is the position for the first constituent of a clause. This position can be filled by different elements, e.g. subject, object or various types of adverbials. The second field is

<sup>66</sup> The adverb *kanske* does exist in Danish but is archaic (only 45 hits in KDK).

called ‘type’, abbreviated as ‘T’. In main clauses, ‘type’ is filled by the finite verb. In subordinate clauses, ‘type’ is filled by a (covert) complementizer. The letter ‘s’ stands for subject position, the capital letter ‘S’ is the object position, ‘V’ is a position for verbs, ‘a’ is the canonical position for sentence adverbs and the capital letter ‘A’ is the field for various kinds of adverbials.

In Table 17a, the structure of a declarative main clause is presented. In Table 17b, the structure of a *wh*-question is illustrated. In yes/no-questions, the finite verb is in ‘type’ and the ‘foundation’ is empty, as in Table 17c. All examples in this section are in Swedish, unless stated otherwise, because Swedish exhibits the most peculiar structures with respect to MON and MAYBE, as will become clear in the remainder of this section.

	F	T	s	a	V	S	A
a.	<i>Olle</i>	<b><i>kommer</i></b>					<i>i morgon.</i>
	Olle	comes					tomorrow
	<i>I morgon</i>	<b><i>kommer</i></b>	<i>Olle</i>				
	Tomorrow	comes	Olle				
b.	<i>När</i>	<b><i>kommer</i></b>	<i>Olle?</i>				
	When	comes	Olle?				
	<b><i>Vem</i></b>	<b><i>har</i></b>	<i>Olle</i>		<i>besökt?</i>		
	Who	has	Olle		visited?		
c.		<b><i>Kommer</i></b>	<i>Olle</i>				<i>i morgon?</i>
		Comes	Olle				tomorrow?
		<b><i>Har</i></b>	<i>Olle</i>		<i>besökt</i>	<i>henne?</i>	
		Has	Olle		visited	her?	

**Table 17.** V2 in the Mainland Scandinavian languages.

In Mainland Scandinavian, declarative main clauses with a subject as the first constituent, as in Table 18a, seem to have the same word order as subordinate clauses, as in Table 18b. The difference in word order between main and subordinate clauses can be made visible by the position of sentence adverbs or a negation marker (Holmberg & Platzack 1995:45). In declarative main clauses the finite verb occurs in second position followed by the negation marker *inte* ‘not’ (Table 18c). In subordinate clauses, the finite verb occurs after the negation marker *inte* (Table 18d).

	F	T	s	a	V	S	A
a.	<i>Olle</i>	<b><i>kommer</i></b>					<i>i dag.</i>
	Olle	comes					today.
b.		<i>att</i>	<i>Olle</i>		<b><i>kommer</i></b>		<i>i dag.</i>
		that	Olle		comes		today.



c.	<i>Olle</i>	<b><i>kommer</i></b>		<b><i>inte</i></b>			<i>i dag.</i>
	Olle	comes		not			today.
d.		<i>att</i>	<i>Olle</i>	<b><i>inte</i></b>	<b><i>kommer</i></b>		<i>i dag.</i>
		that	Olle	not	comes		today.

**Table 18.** Differences in word order between main and subordinate clauses in the Mainland Scandinavian languages.

To illustrate the unusual syntactic behaviour of MON and MAYBE, both V2 and non-V2 clauses with MON and MAYBE are placed and contrasted in Platzack's scheme for main and subordinate clauses.

#### 4.2.2.1 Formal properties of Mainland Scandinavian MAYBE

Adverbs of the MAYBE-type may violate the Verb Second principle (cf. Andréasson 2002; Platzack 1998; Teleman 1974 on Swedish *kanske*). It is only in declarative main clauses that *kanske* may violate the rules of Swedish word order. As the examples in Table 19 show, both V2 (a, b) and non-V2 (c,d) word order are possible in declarative main clauses with *kanske*. Other sentence adverbs cannot occur in non-V2 clauses (e,f), but focus adverbs like for example *också* 'also' or *bara* 'only' can (g, h). The structures in (a, b, d) are also available for Danish *måske* and Norwegian *kanskje*. Note that (c) is a blatant violation of V2. This structure is ungrammatical in Danish and Norwegian but perfectly possible in Swedish.

	<b>F</b>	<b>T</b>	<b>s</b>	<b>a</b>	<b>V</b>	<b>S</b>	<b>A</b>	
a.	<i>Olle</i>	<i>har</i>		<b><i>kanske</i></b>	<i>läst</i>	<i>boken.</i>		V2
	Olle	has		maybe	read	book.the		
b.	<b><i>Kanske</i></b>	<i>har</i>	<i>Olle</i>		<i>läst</i>	<i>boken</i>		V2
	Maybe	has	Olle		read	book.the		
c.	<i>Olle</i>	<b><i>kanske</i></b>			<i>har läst</i>	<i>boken</i>		nonV2
	Olle	maybe			has read	book.the		
d.	<b><i>Kanske</i></b>	<i>(att)</i>	<i>Olle</i>		<i>har läst</i>	<i>boken</i>		nonV2
	Maybe	<i>(that)</i>	Olle		has read	book.the		
e.	<b>*<i>Olle</i></b>	<b><i>förmodligen</i></b>			<i>har läst</i>	<i>boken</i>		
	*Olle	supposedly			has read	book.the		
f.	<b>*<i>Förmodligen</i></b>	<i>(att)</i>	<i>Olle</i>		<i>har läst</i>	<i>boken</i>		
	*Supposedly	<i>(that)</i>	Olle		has read	book.the		
g.	<i>Olle</i>	<b><i>bara läser</i></b>				<i>böcker.</i>		
	Olle	only reads				books.		
h.	<b><i>Också Olle</i></b>	<i>har</i>			<i>läst</i>	<i>boken</i>		
	Also Olle	has			read	book.the		

**Table 19.** Syntactic positions for sentence adverbs in Mainland Scandinavian.

According to Platzack (1998:89; 1999:100) *kanske* is positioned in the *type*-field in all non-V2 structures, but I think this is only true for the structure in (Table 19c). In structure (Table 19d), *kanske* is followed by optional ‘that’ and subordinate word order, which suggests that it is positioned in the *foundation*-field and the (covert) complementizer is positioned in the *type*-field.

In (Table 19c), *kanske* is positioned in the *type*-field, which prevents the finite verb from being placed there. This is reflected in the subordinate word order after *kanske*. Note that it is not possible that both *kanske* and the finite verb are in the *type*-field. A clause like (111) is ungrammatical (Vikner 1995:45).

	<b>F</b>	<b>T</b>	<b>s</b>	<b>a</b>	<b>V</b>	<b>S</b>	<b>A</b>
(111)	* <i>Igår</i>	<i>kanske har</i>	<i>Olle</i>	<i>inte</i>	<i>läst</i>	<i>boken.</i>	
	*Yesterday	maybe has	Olle	not	read	book.the	

The reason why Swedish *kanske* can occupy the position of the finite verb is probably that it has preserved verbal properties. Andréasson (2002:45) remarks that in order to position *kanske* in the *type*-field it has to be defined differently from other sentence adverbs, as a ‘clausal word’ or a word with clause-like properties. To allow both finite verbs and the sentence adverb *kanske* in the *type*-field would be unsatisfactory.

In all syntactic variants, *kanske* functions as sentence adverb, but it in non-V2 structures it is also a ‘clausal word’ because it ultimately derives from a finite verb phrase. Because of this, the subordinating properties of the original finite verb phrase are only retained in the non-V2 structures. The syntactic distributions of Mainland Scandinavian MAYBE will be further discussed and explained in Section 4.4.

#### 4.2.2.2 Formal properties of Mainland Scandinavian MON

The peculiar status of *månne* is comparable to *kanske*, especially the structure in Table 19d, which resembles Table 20a below. As a sentence adverb *månne* takes similar positions as negation markers and adverbs (b). Note that in (b), *månne*, takes a sentence adverbial position, but it still forms a unit with the *wh*-word. This can be shown by the fact that the epistemic sentence adverb *kanske* is not possible in this structure, i.e. *kanske* cannot be the questioned constituent, nor can it convey a speech-act in *wh*-questions. As an epistemic or speech-act adverb *månne* is largely interchangeable with the epistemic or speech-act adverb *kanske*, depending on the contexts, as they are near-synonyms, as in (c). As a speech-act adverb, *månne* may be used as an adjunct, i.e. as a syntactically externalized afterthought, as in (d).

	F	T	s	a	V	S	A
a.	<b>Månne</b>	(om)	han		kommer		i dag?
	MON	(if/whether)	he		comes		today?
b.	<i>Vad</i>	<i>hade</i>	<i>han</i>	<b>månne</b> <b>*kanske</b>		<i>för</i> <i>bakgrund?</i>	
	What	had	he	MON *MAYBE		for background?	
c.	<b>Månne</b>	<i>får</i>	<i>vi</i>		<i>se</i>	[...]	
	MON	get	we		see	[...]	
d.		<i>Fick</i>	<i>hon</i>			<i>Nobelpriset,</i>	<b>månne?</b>
		Got	she			Nobel prize.the,	MON?

**Table 20.** The peculiar syntactic status of Swedish *månne* (1).

A particularly complex structure in Swedish is presented in Table 21 below. Here the status of *månne* is unclear which is reflected by the multitude of possible analyses. Analysis b3 is probably the best option. Structure b1 is ruled out because either a complementizer or the finite verb should be in ‘type’. Here, the finite verb is in the ‘foundation’ which is the position for various types of constituents, except for the finite verb. The structure in b2 is not possible because there cannot be more than one constituent in ‘type’, cf. example (111) on *kanske*. In this particular example, *månne* seems to function as a focus particle that attaches to *Mauritsius*, as shown in b3.

	F	T	s	a	V	S	A
b1.	<i>Är</i>	<b>månne</b>	<i>Mauritsius</i>			<i>ett trevligt ställe</i> [...]	
b.2		<i>Är månne</i>	<i>Mauritsius</i>			<i>ett</i> <i>trevligt</i> <i>ställe [...]</i>	
b.3 <sup>67</sup>		<i>Är</i>	<b>månne</b> <i>Mauritsius</i>			<i>ett</i> <i>trevligt</i> <i>ställe [...]</i>	
	Is	MON	<i>Mauritsius</i>			a pleasant place [...]	

**Table 21.** The peculiar syntactic status of Swedish *månne* (2).

<sup>67</sup> This type of structure is also found with the negative marker *inte* in *Oscar Wilde, är inte han fikus?* ‘Oscar Wild, isn’t he gay?’ or *kanske* as in *Är kanske alla hjärnskadade* ‘Is perhaps everyone brain-damaged?’ Examples from Google.

According to Erteschik-Shir (2010), the Danish example<sup>68</sup> in (112a) is problematic because it cannot be syntactically explained. She proposes that the *wh*-word and MON invert for prosodic reasons: MON + *wh*-word is ruled out as a prosodic constituent whereas *wh*-word + MON is not. She argues that this must be the case because V2 is required in clauses with internal MON, as in (112b).

(112)	F	T	s	a	V	S	A
a	<i>Hvad</i>	<b>mon</b>	<i>de</i>		<i>ville have sagt?</i>		
	What	MON	they		would have said?		
b.	<i>Hvad</i>	<i>ville</i>	<i>de</i>	<b>mon</b>	<i>have sagt?</i>		
	What	would	they	MON	have said?		

I think that the answer to this ‘problematic’ case lies in the history of MON. The structure *wh*-word + MON may be a remnant of an older structure with auxiliary MONNE, as in (113) below.

(113)	F	T	s	a	V	S	A	
	<i>Hvorefter</i>	<b>monne</b>	<i>han</i>		<i>lede?</i>			ODS
	What	MON <sub>AUX</sub>	he		looking for?			

Throughout time, the history of MON became opaque. It gradually lost its canonical position as first constituent and is placed more often in the middle or at the end of a clause. Consequently the word order becomes like *wh*-questions, as illustrated by example (112b). Note that MON in (112b) functions as a sentence adverb, in (112a) it is part of a complex question particle and it retained the tense properties of auxiliary MON.

In sum, both MAYBE and MON may violate V2. The main difference between the two is that MON occurs in various types of interrogatives, whereas MAYBE as an epistemic adverb is ruled out in interrogatives. In case MAYBE occurs in interrogative contexts it is a speech-act adverb. The syntactic distribution of MON differs significantly in the Mainland Scandinavian languages, as will be become clear in the next section.

#### 4.3 Case study IIa: Mainland Scandinavian MON

This section is concerned with a comparative synchronic corpus investigation of the semantic and formal properties of Mainland Scandinavian MON ‘I wonder’, its etymology and development in relation to the lexicalization-grammaticalization-pragmaticalization interface and (inter)subjectification.

Ultimately, MON developed out of a verb into a modal auxiliary, and subsequently into an interrogative/epistemic/speech-act adverb. The present tense of the now deficient auxiliary MONNE ‘may, shall’ (< Old Scandinavian *munu* ‘shall, will’), i.e., Danish *monne*, Norwegian

<sup>68</sup> There are no instances of this structure in the Norwegian and Swedish samples.

*monne* and Swedish *månde*<sup>69</sup>, MON, was reanalyzed as an adverb (i.e. Danish *mon*, Norwegian *mon* and Swedish *månne*).

This development is illustrated in Table 22 by means of the same Danish example sentence in two consecutive periods. The infinitival form *være* ‘be’ that follows MON in (16a) changed into the present tense form *er* ‘is’, as shown in Table 22. In (a) MON is the tense marker of the clause because it is still an auxiliary form, but in (b) MON is an interrogative adverb and tense is expressed by *er*.

	F	T	s	a	V	S	A	
a.		<b>MON</b> <sub>AUX</sub>	<i>det</i>		<i>være</i>	<i>sandt?</i>		ODS
		MON <sub>AUX</sub>	<i>it</i>		<i>be</i>	<i>true?</i>		
b.	<b>MON</b> <sub>ADV</sub>	<i>(om)</i>	<i>det</i>		<i>er</i>	<i>sandt?</i>		ODS
	MON <sub>ADV</sub>	<i>(if)</i>	<i>it</i>		<i>is</i>	<i>true?</i>		

**Table 22.** The reanalysis from auxiliary form to interrogative adverb for MON.

The development from modal auxiliary to adverb does not seem to be unique to MON. For Swedish, the modal verbs *lär*, *torde* and *må*<sup>70</sup> also seem to be on their way of becoming particles (Sundman 1983:328). In the literature, this development is scarcely mentioned as a possible developmental path (e.g. Heine 1993: 77). It is, however, a widespread tendency in Germanic languages (cf. Nuyts 2001) as many adverbs derive from, or contain auxiliary forms, e.g. English *maybe* (< may + be) or Dutch *mogelijk* ‘possibly’ (< *mogen* ‘may’+ *-lijk* ‘-ly’).

Nowadays, the auxiliary form MONNE is archaic and only rarely used in fixed (concessive) expressions. The question particle MON is used in different types of interrogative clauses to express various dimensions of doubt and uncertainty from the point of view of the speaker. Both the auxiliary MONNE and the epistemic/interrogative adverb MON are infrequent linguistic items these days.

This section is organized as follows: Section 4.3.1 is a discussion of the semantic properties of MON, and its ancestor MONNE, in Danish, Norwegian and Swedish. In Section 4.3.2 the sources and method to this case study are presented. Section 4.3.3 presents the results of a small comparative corpus investigation of MON in the Mainland Scandinavian languages. In Section 4.3.4 the status of MON is discussed with regard to the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization as in Table 10 in Chapter 2 and (inter)subjectification. Finally, Section 4.3.5 contains a summary, discussion and conclusions of this case study.

<sup>69</sup> Swedish used to have the forms *monne/monde* and *mon*. The change of *o* > *å* is probably due to influence of the auxiliaries *må*, *måtte* ‘may, must’ (Björkstam 1919:97-143; SAOB).

<sup>70</sup> Note the close resemblance between optative *må* (in (96)) and *måtte* (in (101)) in Chapter 3, and MON in example Table 22. Optative *må* and *måtte* have the structure in (a), but are close to becoming optative markers in first position.

### 4.3.1 Semantic properties of MON and MONNE

This section discusses the semantic properties of MON, and its ancestor MONNE in Mainland Scandinavian in addition to the syntactic peculiarities as discussed in Section 4.2.2.2. Subsection 4.3.1.1 describes the properties of *monne* and *mon* in Danish, in Section 4.3.1.2 the properties of Norwegian *monne* and *mon* are listed, and Section 4.3.1.3 examines the characteristics of Swedish *månde* and *månne*.

#### 4.3.1.1 Danish *monne* and *mon*

Danish *monne* ‘may, shall’ is a modal auxiliary that derives from the Old Danish verb *munu/monu* ‘shall, will’ and is used to express possibility and probability in formal and ironic language use. *Monne* cannot be inflected, is synonymous to *måtte* ‘may, must’ and is followed by infinitival complements, as shown in (114).

- (114) [...] – *abonnenten kan bare ikke få forbindelse med dem,* DDO  
*som **monne** være i den anden ende.*  
‘The subscriber just cannot get connected with those,  
who may be at the other end.’

The interrogative/epistemic adverb *mon* ‘I wonder’ has various meanings. It is used to express uncertainty or doubt in (indirect) questions, as in (115a). In combination with the negation marker *ikke* ‘not’ it may be used to express high probability, as exemplified in (115b). As a short answer it denotes that one does not believe that a certain claim holds, as in (115c). *Mon* may also be used as a polite request, see (115d). The form *mon* may also be part of the set phrase, *mon ikke*, meaning ‘obviously, absolutely’, as represented by (115e).

- (115) a. *Hvilken Bordeaux-vin er **mon** den bedste?* DDO  
‘Which Bordeaux wine is MON the best?’
- b. *De får denne lille jule-idé-bog helt gratis med Hjemmet i denne uge.* DDO  
***Mon** der ikke er en idé eller to, De kan bruge.*  
‘You get the little Christmas idea book entirely for free with this week’s Hjemmet. No doubt you will get an idea or two that are useful.’
- c. *En gang imellem er jeg nødt til at vise mig derhjemme. Ellers glemmer* DDO  
*min kone og børnene, hvordan jeg ser ud. Åh, **mon**?*  
‘Once in a while I have to show up at home. Otherwise my wife and kids will forget what I look like. Oh, really?’

- d. *Fru Engholm! sagde han endelig. **Mon** De kunne skaffe os en liste over alle de ansatte på Epolux?* DDO  
 ‘Fru Engholm! he finally said. Perhaps you could provide us a list of all employees at Epolux?’
- e. *Sagde du ikke, at han er en elendig soldat? – **Mon ikke!** Han kan hverken fægte eller skyde.* DDO  
 ‘Did not you say that he is a bad soldier? - Absolutely! He can neither fight nor shoot.’

#### 4.3.1.2 Norwegian *monne* and *mon*

In Norwegian the form *monne* can either be a full verb or a more or less archaic auxiliary form. As a full verb the different meanings vary between ‘to help/work, be helpful/suitable’ as in (116a) or ‘to grow, increase, progress’ as in (116b). As an auxiliary, the meaning is ‘may, shall’ and *monne*, *mon* (present tense) occurs mainly in concessive constructions (116c,d).

The two variants of *monne* have different etymologies, the full verb *monne* derives from Old Norse *muna* ‘remember’ and auxiliary *monne* originates from Old Norse *munu* ‘shall, will’. The auxiliary has a deficient verbal paradigm that only consists of present tense *mon* and past tense *monne*. The main verb *monne* has a full verbal paradigm.

- (116) a. *Det er noe som **monner**.* BO  
 ‘There is something that works.’
- b. *Det **monner** jevnt.* BO  
 ‘It increases evenly.’
- c. *Jeg reiser, hva hun enn **mon** si.* BO  
 ‘I am going, no matter what she may say.’
- d. *Jeg ønsker dem alt godt, hvor de enn **mon/monne** være.* BO  
 ‘I wish all the best for them, where-ever they may be.’

The adverbial form *mon* is primarily a question particle meaning ‘I would like to know if’, as shown in (117a), it may also occur in combination with the verb *tro* ‘believe, think’, meaning ‘I wonder’, as exemplified in (117b).

- (117) a. ***Mon** hun kommer?* BO  
 ‘I would like to know if she comes.’
- b. ***Mon tro** hva hun sier når hun får se dette?* BO  
 ‘I wonder what she will say if she gets to see this.’

#### 4.3.1.3 Swedish *månde* and *månne*

The deficient auxiliary *månde* derives from the Old Swedish verb *munu* ‘will, shall’. The meaning of *månde* comprises the modal dimensions ‘may, might, can, could’, as in (118a), *månde* occurs mainly in concessive contexts (118b,c). *Månde* may also refer to future possibilities, meaning ‘would, could’, often in literary language use (118d). *Månde* cannot be inflected.

- |          |   |      |
|----------|---|------|
| (118) a. | <i>Vad månde det betyda.</i><br>‘What could it mean.’                       | SAO  |
| b.       | <i>Vem det vara månde.</i><br>‘Who-ever it may be.’                         | SAO  |
| c.       | <i>Vad han än månde säga.</i><br>‘Whatever he may say.’                     | NSSO |
| d.       | <i>Vad månde bliva av detta barn.</i><br>‘What is to become of this child.’ | NSSO |

The interrogative/epistemic adverb *månne*<sup>71</sup> is common in Finland Swedish (Keski-Raasakka 2002) but infrequently used or even considered to be obsolete in Swedish spoken as in Sweden (cf. Teleman et al. 1999; Petersson & Josefsson 2010). Adverbial *månne* may mean ‘I would like to know / I wonder if’, as in (119a). It is often accompanied by a suggested or indicated answer, as shown in (119b), and it may be used as a filler to mitigate a question, as exemplified in (119c). In these contexts *månne* may be interchangeable with *kanske* ‘maybe, perhaps.’

- |          |  |      |
|----------|--|------|
| (119) a. | <i>Månne han kommer ikväll?</i><br>‘I wonder if he comes tonight.’   | SAO  |
| b.       | <i>Månne detta är vad hon har tänkt sig?</i><br>‘I wonder if this is what she expected.’                   | NSSO |
| c.       | <i>Fanns det månne något farligt djur i närheten?</i><br>‘Was there perhaps some dangerous animal around?’ | NSSO |

#### 4.3.2 Method and sources

The Danish data have been selected from *KorpusDK*, the Norwegian data have been collected from *Norsk Aviskorpus* and the Swedish data have been taken from the subcorpora P95-98 of *Språkbankens konkordanser* (Section 1.4). Random samples of 100 instances of MON have been

<sup>71</sup> The form *mån(n)* used to be available as well (SAOB).



taken from the Danish, Norwegian and Swedish corpora. Since MON is an infrequent linguistic item these days, the corpora did not always contain 100 instances of MON.

The data are classified according to the meanings and uses of MON as described in the dictionaries mentioned in Section 4.3.1, i.e. the examples in (115), (117) and (119), which results in six different categories to which MON can be assigned, as shown in (120) below.

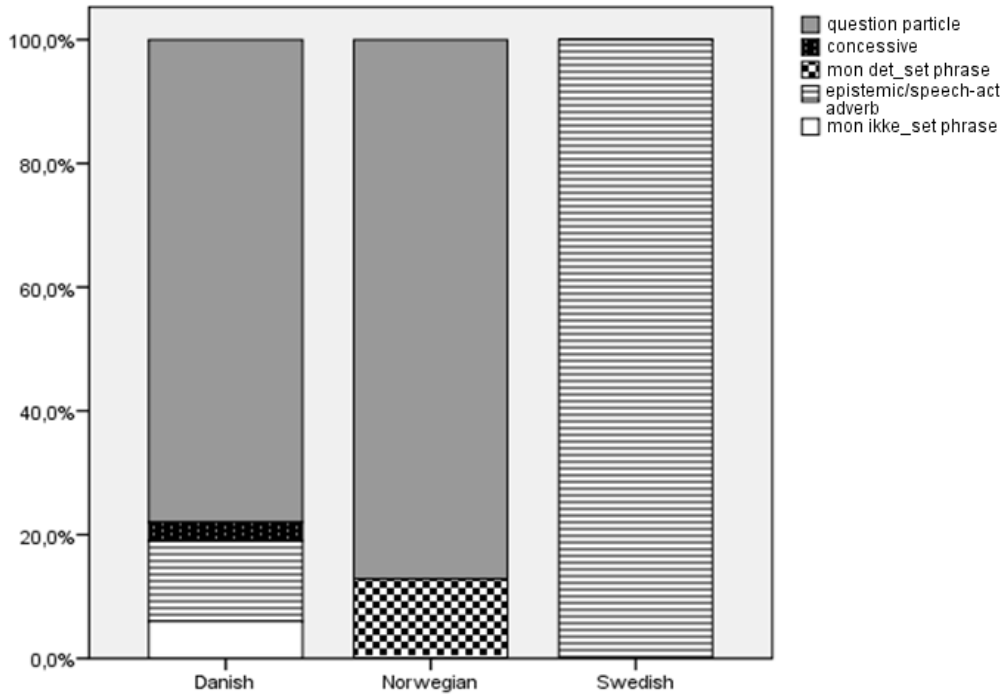
- (120) a. question particle  
b. expression *Mon?*  
c. epistemic/speech-act adverb  
d. expression *Mon tro*  
e. expression *Mon ikke!*  
f. polite request

The distributional properties of Danish *mon*, Norwegian *mon* and Swedish *månne* will be discussed in more detail in the next section.

#### 4.3.3 Results

The distribution of the different meanings and uses of Mainland Scandinavian MON is shown in Figure 7 below. The predefined categories in (120) turned out to be insufficient to classify all of the corpus data. Some categories were not present in the corpus data, but the corpus data also contained meanings and uses that were not mentioned in the dictionaries. Consequently, new categories have been added to the inventory of meanings and uses of MON. See Appendix 3, for an overview of the counts per category in each language.

The distribution of MON is significantly different in Danish, Norwegian and Swedish ( $\chi^2=242.6$ ,  $df=8$ ,  $p<0.01$  and Cramer's  $V=0.678$ ). The high value of Cramer's  $V$  indicates a strong association between language and meaning/use. It points to a language-specific development for Swedish *månne*, the distribution of which is clearly different from Danish and Norwegian *mon*. In Norwegian and Danish *mon* is mainly used as a question particle or it occurs in set phrases. In Swedish the use and meaning of *månne* is similar to that of the epistemic/speech-act adverb 'maybe'.



**Figure 7.** Bar chart of the different meanings and uses of MON.

The different meanings of Danish *mon*, Norwegian *mon* and Swedish *månne* will be discussed in more detail by means of corpus examples in 4.3.3.1-3 below. The overall picture that arises from the distributional properties of MON is summarized in 4.3.3.5.

#### 4.3.3.1 Danish *mon*

*KorpusDK* contains 3445 hits for *mon*. The sample of 100 occurrences of *mon* consists of 78 instances of a question particle, as in (121a). Question particles may be accompanied by a *wh*-element, as shown in (121b), or the negation marker *ikke* ‘not’, as exemplified in (121c). Thirteen times *mon* is used as an epistemic or speech-act adverb, see (121d). There are six instances of the set phrase *mon ikke*, as in (121e), and three excluded cases in which *mon* was not an interrogative adverb.

- (121) a. *Mon han også var nervøs?*  
 ‘I wonder if he was nervous as well.’
- b. *Hvor mon de var taget hen?*  
 ‘Where MON did they go?’

- c. **Mon ikke** *det snart er på tide at udskrive en idé-konkurrence om den dokumentarfilm om Lars von Triers arbejdsmetode, der endnu ikke er blevet lavet.*  
'Isn't it about time to hold a competition for a documentary movie about the working method of Lars von Trier, which has not been made yet.'
- d. *Betyder det mon noget?*  
'Does it perhaps mean anything?'
- e. *"Kan du lide den? spørger far. Ja, mon ikke!"*  
'Do you like this one?', dad asks. Yes, for sure!'

#### 4.3.3.2 Norwegian *mon*

*Norsk Aviskorpus* contains 1445 hits for *mon*. The sample of 100 occurrences of *mon* consists of 75 question particles, as represented by (122a), which may be accompanied by the negative marker *ikke* 'not', as in (122b). Of these, there are 64 instances of the combination *mon tro*, as in (122c). The set phrase *mon det* 'really?/so be it' occurs eleven times in the sample, see (122d).

- (122) a. **Mon om** *noen av moteukene kommer til å fokusere på det?*  
'I wonder if some of the fashion weeks will focus on it.'
- b. **Mon ikke** *det bare går oppover etter dette?*  
'I wonder if it is not just on the way up after this?'
- c. **Mon tro om** *den smaker godt.*  
'I wonder if it tastes good.'
- d. **Mon det?** *Tønne visste ikke.*  
'Really? Tønne did not know.'

#### 4.3.3.3 Swedish *månne*

*Språkbankens konkordanser* contains 331 instances of *månne*, the sample from P95-98 consists of 78 occurrences. The exact meaning/function of *månne* is hard to determine. Its meaning is overlapping between 'I wonder, could it be, maybe', as in (123a). *Månne* may occur in various positions for sentence adverbs, see (123b). This suggests that Swedish *månne* has gradually become a sentence adverb that is largely synonymous to the epistemic adverb *kanske* and which developed speech-act uses. Note that Verb Second (V2) word order is possible for clauses with *månne* as first constituent, as exemplified in (123c). This is not (yet) the case for Danish and Norwegian.

- (123) a. *Victoria, **månne**?*  
‘Victoria, perhaps?’
- b. *Är det **månne** Pettson själv som spelar?*  
‘Is it perhaps Pettson himself who plays?’
- c. ***Månne** får vi se ännu fler bankchefer gråta ut framöver.*  
‘Maybe we get to see even more bankers break down in the future.’

#### 4.3.3.4 Overall results

The results this case study on Mainland Scandinavian MON highlights considerable differences, both formally and semantically, with respect to its status and development. As shown in example (121)-(123), there are substantial differences when it comes to available syntactic structures for clauses containing the interrogative adverb MON in Mainland Scandinavian. All in all, it can be observed that Mainland Scandinavian MON exhibits clearly language specific patterns and idiosyncrasies.

In Danish, MON often co-occurs with a *wh*-word or the negation marker *ikke* ‘not’, overt conjunctions are not attested in the sample, *mon* may also function as a sentence adverb. The most frequent structures with MON are shown in Table 23 below.<sup>72</sup>

	F	T	s	a	V	S	A
a.	<i>Hvad</i>	<b>mon</b>	<i>de</i>		<i>ville have sagt?</i>		
	What	MON	they		would have said?		
b.	<i>Hvornår</i>	<i>lærte</i>	<i>hun</i>	<b>mon</b>	<i>at disponere</i>	<i>sin tid</i>	<i>fornuftigt?</i>
	When	learned	she	MON	to manage	her time	in a sensible way?
c.	<b>Mon</b>	<i>(om)</i>	<i>det</i>		<i>er</i>	<i>prisen</i>	<i>værd, social-demokrater?</i>
	MON	(if)	it		is	price.the	worth, social democrates?

<sup>72</sup> The conjunctions (placed in brackets) are not present in the actual corpus examples (the presence of *om* sounds odd and there are no examples of *Mon om* in KDK), but they seem to be implicitly present as can be inferred from the visualized syntactic structures in the sentence scheme. That is, some (covert) element must be located in T, otherwise the finite verb would have been in T instead of in V. But see Table 24 on Norwegian *mon* which generally requires the presence of *om*.

d.	<b>Mon ikke</b>	(om)	de		finder på	noget nyt	
	MON not	(whether)	they		come across	something new	
e.		Betyder	det	<b>mon</b>		noget?	
		Means	it	MON		something?	

**Table 23.** The syntactic distribution of Danish *mon*.

In Norwegian the interrogative adverb MON mainly occurs together with the cognitive verb TRO, which is optionally followed by an overt complementizer. The most frequent syntactic structures with *mon* (*tro*) in Norwegian are summarized in Table 24 below.

	<b>F</b>	<b>T</b>	<b>s</b>	<b>a</b>	<b>V</b>	<b>S</b>	<b>A</b>
a.	<b>Mon</b>	(om)	det	likevel ikke	er	denne tilværelsens kortvarige letthett,	som [...]
	MON	(if)	it	nonethe- less not	is	this life's short-lived ease,	that [...]
b.	<b>Mon</b>	om	de	ikke	var litt forelsket i henne	alle sammen,	likevel?
	MON	if	they	not	were little in love with here	all together	none- theless.
c.	<b>Mon tro</b>	om	de		er	like glad for dette	idag?
	I wonder	if	they		are	equally happy about this	today?
d.	<b>Mon tro hva</b> <sup>73</sup>	(som)	han		tenkte på		den gangen?
	I wonder what		he		thought about		this time. the.

<sup>73</sup> Platzack (1998:93) notes that, for Swedish, subordinate interrogative clauses with overt *som* 'as' in which the questioned constituent is not the subject, become more acceptable the more complex the questioned constituent is. Compare ?*Jag undrar vad som han har köpt* 'I wonder what he has bought' and *Jag undrar vad för sorts bok som han har köpt* 'I wonder what kind of book he bought'. Note that *montro* is interchangeable with, and has the same status, as the phrase *jag undrar*.

e.	<i>Hva</i>	<i>ble</i>	<i>disse pengene</i>		<i>brukt til,</i>		<b><i>mon tro?</i></b>
	What	was	this money		used for		, I wonder?

**Table 24.** The syntactic distribution of Norwegian *mon*.

Swedish *månne* functions primarily as an epistemic sentence adverb or speech act adverb in interrogative clauses, as illustrated in Table 25. Because of this, V2 word order is possible in clauses with *månne* as the first constituent. This is not possible for Danish and Norwegian as *mon* is primarily an interrogative adverb in these languages, not a sentence adverb.

	<b>F</b>	<b>T</b>	<b>s</b>	<b>a</b>	<b>V</b>	<b>S</b>	<b>A</b>
a.	<b><i>Månne</i></b>	<i>(om/att)</i>	<i>man</i>	<i>då</i>	<i>kommer fram till</i>	<i>att pengarna kunde användas klokare?</i>	
	MON	(if/that)	one	then	realizes	that the money could be used more cleverly?	
b.	<b><i>Månne</i></b>	<i>kommer</i>		<i>då</i>		<i>nya överraskningar?</i>	
	Maybe	comes		then		new surprises?	
c.		<i>Är</i>	<b><i>månne</i></b> <i>arvodena</i>			<i>högre,</i>	<i>då?</i>
		Are	perhaps fees.the			higher,	Then?
d.		<i>Är</i>	<i>det</i>	<b><i>månne</i></b>		<i>den rätte Gerner?</i>	
		Is	it	perhaps		the right Gerner	
e.		<i>Fick</i>	<i>hon</i>			<i>Nobelpriset,</i>	<b><i>månne?</i></b>
		Got	she			Nobelprize. the,	perhaps?

**Table 25.** The syntactic distribution of Swedish *månne*.

The general picture is that Danish, Norwegian and Swedish developed their own idiosyncrasies and syntactic structures for MON. Whenever MON occurs as the second constituent in a clause, it retains traces of the auxiliary form MONNE. As an interrogative adverb it has subordinating properties because a complementizer like IF or a *wh*-word is (covertly present).<sup>74</sup> In V2 structures MON has lost the subordinating properties of the interrogative adverb and functions as an epistemic sentence adverb or speech-act adverb.

#### 4.3.4 Mainland Scandinavian MON: a case of grammaticalization?

As we have seen in the previous sections, the development of Mainland Scandinavian MON is a development from minor category (modal auxiliary) to minor category (interrogative adverb), and subsequent shifts within the broad domain of adverbs (epistemic/speech-act adverb). In what follows, I will discuss and analyze the development of Mainland Scandinavian MON in relation to the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization and (inter)subjectification.

First the essential mechanisms in lexicalization, grammaticalization and pragmaticalization will be discussed in subsection 4.3.4.1 then the accompanying primitive changes will be examined in subsection 4.3.4.2, the possible side effects are elaborated on in subsection 4.3.4.3, and finally, in 4.3.4.4 I will comment on the linguistic status (i.e. lexical, grammatical or communicative) and degrees of lexicalization, grammaticalization or pragmaticalization, and (inter)subjectification for Mainland Scandinavian MON.

##### 4.3.4.1 Mechanisms in the development of MON

As defined in the present study, lexicalization, grammaticalization and pragmaticalization are essentially conceived of as the result of formal reanalysis and semantic reinterpretation. With regard to MON, the scheme looks as in Table 26. The cells that apply to MON are shaded.

The first step in the development of MON, is the reanalysis from the present-tense auxiliary form MONNE into the interrogative adverb MON, as illustrated in Table 22 in Section 4.3. Once the adverbial status was established, MON continued to develop into an epistemic sentence adverb and a speech-act adverb which may be syntactically externalized, cf. the examples in Section 4.2.2.

Semantically, there are contiguous, metonymic relations between the various interrogative, epistemic and speech-act meanings of MON (cf. the examples in (120)), these are all nuances in the domain of doubt and uncertainty. An in depth historical survey may reveal the small steps that eventually led to these changes in meaning as well as the bridging-contexts for these semantic changes.

---

<sup>74</sup> See also the synonymous expressions GADD VITE followed by *om/wh*-word and JEG UNDRER followed by *om/wh*-word. That is, Danish *gad vide om/wh-word*, Norwegian *gadd vite om/wh-word* 'I would like to know if/wh-word' and Danish *jeg undrer om/wh-word*, Norwegian *jeg undrer (på) om/wh-word*, Swedish *jag undrar om/wh-word* 'I wonder if/wh-word'.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	±
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-
<b>reinterpretation</b>					
<b>-metaphor/metonymy</b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

**Table 26.** Mechanisms in the development of MON.

#### 4.3.4.2 Primitive changes in the development of MON

Formal reanalysis and semantic reinterpretation are accompanied by a subset of correlated primitive changes. The primitive changes that apply to MON are represented by the shaded cells in Table 27.

At the level of phonology/phonetics, there is no loss of phonological/phonetic substance in the development of Norwegian and Danish *mon*, at least not in the written language. For Swedish there has been assimilation in the change from *mānde<sub>v</sub>* to *mānne<sub>adv</sub>*. Likewise, there is no change in morphological compositionality. Since MON is a monomorphemic item, and not a compositional form, its internal structure has not been affected by reduction or univerbation. There is loss of morphosyntactic properties due to the shift from auxiliary to adverb. MON no longer has the ability to express tense. The inflectional properties of verbs had already been lost in the course of auxiliation of MONNE.

The reanalysis from auxiliary form to interrogative adverb results in increased syntactic variability and autonomy. The auxiliary form, MON is dependent on the main verb and has a fixed syntactic position. As an interrogative adverb MON becomes more flexible because it may occur in various types of interrogative clauses. At first the canonical position was clause-initial followed by subordinate word order, later on it could occur in other adverbial positions as well. As an epistemic adverb MON is even more flexible in that it does not form a unit any longer with a *wh*-element or interrogative structures in general. The speech-act adverb MON may be an adjunct, in which case it is a syntactically externalized item.



As regards semantic changes, it can be observed that MON has a much bleached meaning. It does not denote referential meaning but different kinds of interrogative meaning. It developed different nuances in the realm of uncertainty and doubt. The exact meaning of MON is, even in context, hard to determine (and to translate). In general, the presence of MON adds a dimension of personal doubt to interrogatives.

In all its functions (i.e. interrogative, epistemic and speech-act adverb) MON is clearly a (inter)subjective item. It denotes uncertainty or doubt from the point of view of the speaker in interrogative contexts. In the rise of epistemic and speech-act meanings there has been a shift from syntactic to speaking subject compared to the interrogative adverb and auxiliary form. The role of (inter)subjectification in the development of MON will be discussed in more detail in subsection 4.3.4.4.

<b>ii. Primitive changes</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>-phonology/phonetics</b>		
° loss of phonological/phonetic substance	(+)	(+)
<b>-morphology</b>		
° loss of morphological compositionality	(+)	(+)
° loss of morphosyntactic properties	(+)	(+)
<b>-syntax</b>		
° loss of syntactic variability	+	-
° loss of syntactic autonomy	+	-
<b>-semantics</b>		
° loss of semantic substance	+	+
° loss of semantic compositionality	(+)	(+)
<b>-discourse/pragmatics</b>		
° subjectification	(+)	+
° intersubjectification	(+)	+

**Table 27.** Primitive changes in the development of MON.

#### 4.3.4.3 Side effects in the development of MON

The side effects of formal reanalysis and semantic reinterpretation and their accompanying primitive changes can be used as a diagnostic to identify potential cases of lexicalization, grammaticalization or pragmaticalization. As regards MON, the side effects that are noticeable in its development are represented by the shaded cells in Table 28 below.

*Paradigmaticization*, generally relates to *productivity* and *frequency*, but they are not that obviously correlated in the development of MON. In a general sense, MON enters the adverbial paradigm, but is restricted to various interrogative contexts, and is a highly infrequent linguistic item. *Obligatorification* clearly does not apply to MON as it is a formulaic, optional element in interrogative clauses. Likewise *condensation* does not occur because the structural scope of MON

is not reduced and neither does it become more dependent upon the *wh*-element or interrogative structure, rather the opposite. To my knowledge, at least in the Germanic languages, there are no attestations of this specific kind of question particle.

With respect to *layering*, we see that the present tense of MONNE was reanalysed as the question particle MON. The question particle MON could in turn be used to express various dimensions of doubt. It may, inter alia, be an interrogative marker, epistemic sentence adverb, speech-act adverb or it may occur in set phrases. This is illustrated by the different meanings and uses (=layers) in (120) and Figure 7.

As regards *divergence* it can be observed that the source to MON, the auxiliary form MONNE, continued to exist besides the new adverbial form. Auxiliary MONNE came to be fixed in expressions with future reference and in concessive contexts. In combination with the verb *tro* ‘believe, think’, the synonymous adverb MONTRO<sup>75</sup>, developed out of phrases such as Danish *mon (jeg/man) skal tro at* or Swedish *månne jag/man skall tro (att)* meaning ‘MON<sub>ADV</sub> I/one will believe that’ (ODS, DDO, SAOB). In Norwegian, the expression *mon tro* is mostly written as two words, the unverbated adverb *montro* occurs only 15 times in *Norsk Aviskorpus*, cf. the overview in Table 16.<sup>76</sup> The short form *tro* (< *mon tro*, *skal tro*), may also be used as question particle as in (124) below.

- (124) a. ***Tro om det blir regn?*** BO  
           ‘I wonder if there will be rain.’
- b. ***Blir det regn, tro?*** BO  
           ‘Will it be raining, I wonder.’

*Specialization* applies to the phenomenon that similar and simultaneously existing expressions are reduced to one major expression. That MONNE now is a deficient auxiliary form of which the derived adverbial form MON is more frequent these days, might be due to the fact that there were other modal auxiliaries that could express the same kind of meanings. In Old Scandinavian MONNE was a frequently used verb (Birkmann 1987; Björkstam 1919), but it seems to have been gradually displaced by other modals with similar meanings. Hence, there might have been competition in the modal domain for the expression of possibility and probability in such a way that other modals (e.g. *må*, *måtte kunne*, *skulle*, *ville*) took over the more general possibility meanings and MONNE and MON developed specific interrogative meanings. A more in depth historical survey of the development of MONNE is needed in order to find out why it became obsolete.

<sup>75</sup> i.e., Danish *monstro*, Norwegian *montro* and Swedish *måntro* (but see also the synonymous adverbs *måtro*, *kanthro*).

<sup>76</sup> Danish *mon* is, according to Erteschik-Shir (2010), a short version of *monstro*. This presupposes that the adverb *monstro* was attested earlier than the interrogative adverb *mon*. I think a parallel development of MON and MONTRO is more likely.

*Persistence* relates to the observation that a linguistic item or construction retains traces of the linguistic item or construction from which it emerged. The interrogative adverb MON retained properties of auxiliary MONNE when it is the second constituent in a clause. As a sentence adverb in the foundation of a clause it retained properties of the interrogative adverb because a (c)overt complementizer is present in these structures.

<b>iii. Side effects of change</b>	<b>Gzn2</b>	<b>Pgzn</b>
-paradigmaticization	+	(+)
-obligatorification	(+)	-
-condensation	+	-
-layering/divergence/specialization/persistence	+	+
-productivity	+	+
-frequency	+	+
-typological generality	(+)	(+)

**Table 28.** Side effects in the development of MON.

#### 4.3.4.4 Linguistic status and degrees of X-ization for MON

As defined in the present study, lexicalization results in a linguistic item belonging to a major category, with referential meaning, primary status, and which may convey the main point of linguistic message, grammaticalization leads to a linguistic item belonging to a minor category, with relational meaning, secondary status, and which regulates grammatical structure, and pragmaticalization results in a linguistic item with communicative meaning, extra-propositional status, and which organizes discourse structure.

With regard to MON, the overall result of the changes is an adverb with interrogative, epistemic and speech-act meanings. The patterns that arise from Table 26-Table 28 show that its development has most properties in common with secondary grammaticalization and pragmaticalization. Since there is no hierarchical reanalysis, the development of the interrogative, epistemic and speech-act adverb are an instance of secondary grammaticalization accompanied by (inter)subjectification at the textual level, as shown in Table 29.

<b>(Inter)Subjectification</b>	<b>MON</b>
<b>I. subjectification</b> [speaker perspective, attitude and judgment]	+
-textual level [text-construction + meta-linguistic meanings] interrogative, epistemic meanings	
<b>III. intersubjectification</b> [interaction with interlocutor]	+
-textual level interrogative, speech-act meanings	

**Table 29.** Types of subjectification in the development of MON.

In the overall development, Swedish has progressed most in that *månne* is primarily an epistemic/speech-act adverb. This is also reflected by the fact that V2 structures with *månne* in the foundation and the finite verb in type are only possible in Swedish. In Danish, sentence-internal *mon* is possible in V2 clauses. In Norwegian, *mon* only occurs in non-V2 structures. With respect to V2, MON has a lot in common with the epistemic/speech-act adverb MAYBE, as will become clear in the next section.

#### 4.4 Case study IIb: Mainland Scandinavian MAYBE

This section is concerned with a comparative synchronic corpus investigation of the semantic and formal properties of Mainland Scandinavian MAYBE, its etymology and development in relation to the lexicalization-grammaticalization-pragmaticalization interface and (inter)subjectification.

Ultimately, these epistemic adverbs are univerbations of a modal form ‘can/may’ and a main verb ‘happen.’ Various etymologies have been proposed, but the true origin of *kanske*, *kanskje* and *måske* is still uncertain. The intriguing question is whether they are loan translation or part of a cross-linguistically regular grammaticalization path. Both possibilities have been proposed with regard to the development of Swedish *kanske*.

Swedish *kanske* may be a loan translation of Middle Low German *mach schên*, Latin *forsitan* or similar verb phrases in other European languages. It may also be the result of a cross-linguistically regular grammaticalization pathway, that is, the univerbation of a modal form ‘can’ or ‘may’ and a main verb meaning ‘happen’ or ‘be’ may have been a process of productive word formation in various languages. Both scenarios account for the existence of similar epistemic adverbs in European languages. Even though the influence of language contact cannot be excluded, I consider a cross-linguistically regular grammaticalization pathway to be the most plausible etymology for *kanske*.

First of all, the development of Swedish *kanske* shows that the verb phrase was variable, i.e., both the modal forms and the main verb were interchangeable with synonymous forms, respectively *må* ‘may’ or *tör* ‘may’ and *hända* ‘happen’ (cf. the synonymous adverbs *kanhända*, *måhända*, *törhända*). Secondly, both *kunna* ‘can / may’ and *ske* ‘happen’ already existed in Old Swedish (Andréasson 2002:26) which enables univerbation of *kan* and *ske* without interference of the Middle Low German verb phrase *mach schên* or similar verb phrases in other European languages. Since the collocation was variable (cf. univerbations with other modal forms or the verb *hända*) and Swedish *kanske* and Norwegian *kanskje* are not literal translations of *mach schên*, I consider a loan translation not to be a plausible etymology for Swedish *kanske* and Norwegian *kanskje*.

For Danish *måske* on the other hand, the link to Middle Low German *mach schên* is more plausible, because this adverb contains the modal verb *må*. It could be that Danish has retained

*må* in its older meaning ‘can.’<sup>77</sup> But it could also be that Danish is influenced by West Germanic languages, where the etymological counterparts of *må* express deontic and/or epistemic possibility (cf. German *mögen* ‘may’, Dutch *mogen* ‘may’, English *may*, but also Danish *måtte* can mean ‘may’) rather than deontic and/or epistemic necessity like in the other North Germanic languages (ODS).

Wessén (1967:15)<sup>78</sup> distinguishes different stages in the development of Swedish *kanske* from verb phrase to epistemic adverb, as summarized in (125).<sup>79</sup> At stage I the formal subject *det* ‘it’ is more or less obligatory, at stage II the formal subject (*det*) is no longer part of the verb phrase, but the complementizer *att* ‘that’ cannot be omitted, at stage III the complementizer *att* can be omitted, at stage IV inversion of subject and finite verb is possible, hence V2 clauses, at stage V *kanske* can occur clause-internally. Nowadays *kanske* can occur in positions for sentence adverbs (adverbial *kanske* in V2 clauses (IV,V) and in positions where it is followed by subordinate word order (subordinating *kanske* in non-V2 clauses), (II, III). Combinations of a modal auxiliary and a main verb with referential subject, as in (I), are also still possible.

(125) I.	<b>(Det)</b>	<b>kan</b>	<b>ske</b>	<b>att</b>	<i>han</i>	<i>kommer</i>	<i>redan</i>	<i>idag.</i>
	It	may	happen	that	he	comes	already	today.
II.	<b>Kanske / kan ske</b>			<b>att</b>	<i>han</i>	<i>kommer</i>	<i>redan</i>	<i>idag.</i>
	Maybe / may be			that	he	comes	already	today.
III.	<b>Kanske</b>		<i>han</i>	<i>kommer</i>	<i>redan</i>	<i>idag.</i>		
	Maybe		he	comes	already	today.		
IV.	<b>Kanske</b>		<i>kommer</i>	<i>han</i>	<i>redan</i>	<i>idag.</i>		
	Maybe		comes	he	already	today.		
V.	<i>Han</i>	<i>kommer</i>	<b>kanske</b>	<i>redan</i>	<i>idag.</i>			
	He	comes	maybe	already	today.			

In late 18<sup>th</sup> century Swedish, a new clause type emerges that violates the V2 principle. In these clauses, *kanske* is the second constituent followed by subordinate word order, as in (126) and

<sup>77</sup> Note that *måske* used to be an epistemic adverb in Swedish as well (SAOB) and that the form *må* is retained in the synonymous adverb *måhända* which is now archaic. Moreover, *kanske* still exists in Danish but is archaic, cf. footnote 64.

<sup>78</sup> For the sake of clarity, the developmental stages are exemplified with Modern Swedish equivalents of the older stages.

<sup>79</sup> This developmental scenario for *kanske* by Wessén (1967) resembles Thompson & Mulac (1991) developmental path for *I think*, as both *maybe* and *I think* may occur in complementing and parenthetical patterns. I will return to this point in Chapter 5.

example Table 19c in section 4.2.2. Note that this clause type is not mentioned in Wessén's description of the development of *kanske* in (125) above.

(126)	<i>Han</i>	<i>kanske</i>	<i>kommer</i>	<i>redan</i>	<i>idag.</i>
	He	maybe	comes	already	today.

The first occurrences of this clause type are found in the work of the Swedish poet and composer Carl Michael Bellman (1740-1795). The use of this clause type strongly increased in the work of the Swedish writer and dramatist August Strindberg (1849-1912). Literary language may be freer and allow for more syntactic variation than ordinary language use, which may have facilitated the rise of this clause type. At the end of the 20th century this structure is well established in the Swedish language. For more details on the etymology and historical development of *kanske*, see Andréasson (2002), Norde, Rawoens & Beijering 2012).

This section is organized as follows: Section 4.4.1 is a discussion of the semantic properties of MAYBE in Danish, Norwegian and Swedish. In Section 4.4.2 the sources and method to this case study are presented. Section 4.4.3 presents the results of a comparative corpus investigation of MAYBE in the Mainland Scandinavian languages. In Section 4.4.4 the development of MAYBE will be discussed with respect to the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization as in Table 10 in Chapter 2 and (inter)subjectification. Finally, Section 4.4.5 contains a summary, discussion and conclusions.

#### 4.4.1 Semantic properties of MAYBE

This section discusses the properties of MAYBE in Mainland Scandinavian as described in dictionaries (DDO, BO, NSSO), in addition to the syntactic properties that have been mentioned in the preface to these case studies in 4.2.2. Subsection 4.4.1.1 describes the properties of Danish *måske*, in subsection 4.4.1.2 the properties of Norwegian *kanskje* are listed, and subsection 4.4.1.3 examines the characteristics of Swedish *kanske*. The dictionaries provide only very brief descriptions and synonyms, but the main point is that MAYBE may either be used as an epistemic sentence adverb or as a speech-act adverb.

##### 4.4.1.1 Danish *måske*

The Danish dictionary notes that *måske* is used to express possibility or probability, as in (127).

- (127) *Måske kommer han slet ikke hjem. Det kan være, han bliver hos Pia i nat.* DDO  
 ‘Maybe he won’t be coming home. It may be, he stays with Pia tonight.’

In addition, *måske* may convey doubt or uncertainty, as in (128).

- (128) a. *Du glemte osten! Men du vil måske ikke have den?* DDO  
 ‘You forgot the cheese! But perhaps you don’t want to have it?’

- b. *Bogens længste og måske mest fascinerende novelle er “Skabet.”* DDO  
 ‘The book’s longest and perhaps most fascinating novel is “The closet.”’

Note that there is a distinction between epistemic and speech-act use of the adverb, and that *måske* may have a modifying function as well, as in (128b).

#### 4.4.1.2 Norwegian *kanskje*

The Bokmål dictionary only states two meanings/uses of *kanskje*. One meaning is very literal ‘can/may happen’, as in (129a), the other meaning/use denotes ‘confirmation/assurance’, as in (129b).

- (129) a. *De kommer kanskje i morgen* BO  
 ‘Maybe they come tomorrow.’
- b. *Fikk jeg ikke rett, kanskje?* BO  
 ‘Wasn’t I right, perhaps?’

Also here, there is a main distinction between epistemic and speech-act use of the adverb *kanskje*.

#### 4.4.1.3 Swedish *kanske*

Norstedts big Swedish dictionary states that *kanske* denotes ‘a certain possibility’ with respect to, amongst others, happenings and properties. It provides a number of linguistic contexts in which *kanske* may occur, see (130), but does not comment any further on the function or nuances in the meaning that *kanske* may convey.

- (130) a. *Hon skulle kanske komma men hon visste inte säkert.* NSSO  
 ‘She was perhaps going to come but she wasn’t sure.’
- b. *Du har kanske hört att sammanträdet är inställt.* NSSO  
 ‘Perhaps you’ve heard that the meeting is cancelled.’
- c. *Kanske berodde det hela på ett misstag.* NSSO  
 ‘Maybe it is all due to a mistake.’
- d. *Han är kanske inte världens bästa lärare.* NSSO  
 ‘He is maybe not the world’s best teacher.’
- e. *Ett kanske orättvist påstående.* NSSO  
 ‘A maybe unjust assertion.’

Here, the speech-act use is not explicitly mentioned, but present in (130b). Note that *kanske* may take on different functions, e.g. epistemic sentence adverb (130a,c,d) or modifying adverb in (130e).

#### 4.4.2 Sources and method

The comparative corpus investigation of MAYBE focuses on its syntactic distribution and degree of adverbialization, i.e. the degree to which MAYBE has become a full-fledged sentence adverb. For Swedish, the results of Andréasson's (2002) study on the number of syntactic positions for *kanske* will be used. For Danish *måske* and Norwegian *kanskje* I will carry out replication studies after Andréasson (2002), in order to obtain comparable data for these languages.

Andréasson's study is based on a random sample of 836<sup>80</sup> sentences, which have been taken from *Språkbankens konkordanser*. Andréasson used the subcorpora *Press 98* (P98) and *Bonniersromaner II* (RII). P98 contains texts from Swedish newspapers. RII is a corpus that contains 60 Swedish novels that are published by *Bonniers* publishing company in the years 1980 and 1981. The Norwegian and Danish random samples consist of 1000 sentences containing *kanskje* and *måske*, which have been selected from the modern online corpora *KorpusDK* and *Norsk Aviskorpus* (Section 1.4). The corpora are comparable in the sense that they all contain newspaper texts, but Andréasson's data and *KorpusDK* contain literary texts as well, which makes the Norwegian corpus less comparable to the Danish and Swedish corpora.

The method for the corpus investigation of Danish and Norwegian MAYBE is similar to Andréasson's (2002) study of Swedish *kanske*. Only declarative main clauses are considered because it is only in these structures that *kanske* may violate the V2 principle. Interrogative clauses, subordinate clauses, clauses without finite verb, fragmented clauses, clauses in which *kanske* does not function as sentence adverb (but as a modifying adverb) and combinations of a modal auxiliary and a main verb with referential subject (*det kan ske att* 'it can / may happen') have been excluded.<sup>81</sup>

The data are analysed and classified according to three basic clause types. The first clause type (henceforth clause type I) is a non-V2 clause with *kanske* as its first constituent followed by a clause which has subordinate word order, as in (131a). In these clauses *kanske* has the subordinating properties of a verb phrase because it is followed by the (covert) complementizer *att* 'that' and subordinate word order. It is possible to paraphrase *kanske* as the verb phrase *det kan ske att*, as in (131b), but it is not possible to substitute *kanske* by synonymous adverbs, for example *sannolikt* 'probably', see (131c).

- (131) a.     ***Kanske***            (*att*)   *han*    *inte*   *kommer*    *idag*.  
               Maybe            (that) he     not     comes     today.

<sup>80</sup> Every fifteenth occurrence of *kanske* in the subcorpora RII and P98 is part of her sample, which results in a total of 836 instances of *kanske*.

<sup>81</sup> These selective restrictions have turned out to be unfortunate for the purpose of the present study, because Andréasson's focus is on the syntactic distribution only, not on the interface between lexicon, grammar and discourse.



- b. *Det kan ske att han inte kommer idag.*  
 It may happen that he not comes today.
- c. *\*Sannolikt (att) han inte kommer idag.*  
*\*Probably (that) he not comes today.*

The two subtypes of clause type I are KS (132a) and KaS (132b) clauses. The examples are taken from Andréasson (2002).

- (132) a. *Kanske doktorn var ett slags vampyr som* **KS**  
 Maybe doctor.the was a sort vampire that  
*kom om natten och sög blod ur den döende.*  
 came at night.the and sucked blood out the dying.  
 ‘Maybe the doctor was some kind of vampire who came at night and sucked blood out of the dying person.’
- b. *Kanske att Mjök-Moelwyn hade ögonen på* **KaS**  
 Maybe that Mjök-Moelwyn had eyes.the on  
*henne.*  
 her.  
 ‘Maybe Mjök-Moelwyn had an eye on her.’

The second clause type (henceforth clause type II) is a V2 clause in which *kanske* can be any constituent except the second one, as in (133a,b). In V2 clauses *kanske* occurs in canonical positions for sentence adverbs. It is not possible to paraphrase *kanske* as the verb phrase *det kan ske att*, as in (133c), but *kanske* can be substituted by synonymous adverbs, for example *sannolikt* ‘probably’, as in (133d).

- (133) a. *Han kommer kanske inte idag.*  
 He comes maybe not today.
- b. *Kanske kommer han inte idag.*  
 Maybe comes he not today.
- c. *\*Han kommer det kan ske att inte idag.*  
*\*He comes it may happen that not today.*
- d. *Han kommer sannolikt inte idag.*  
 He comes probably not today.

The second clause type comprises different types of V2 clauses. These can have a subject, object/predicate or various types of adverbs (including MAYBE) or adverbials as their first constituent. Clause type II has four different subtypes, SF (134a), KF (134b), OF (134c) and AF (134d) clauses.<sup>82</sup>

- (134) a. *Nåja, några små fel var kanske acceptabla* **SF**  
 Well, some little mistakes were maybe acceptable  
*efter ett liv där det mesta varit så rätt.*  
 after a life there the most been so correct.  
 ‘Well, some little mistakes were maybe acceptable after a life in which most things had been so correct.’

- b. *Kanske begravde han på sätt och vis* **KF**  
 Maybe buried he on way and manner  
*också sin egen ungdom där i Kungsträdgården.*  
 also his own youth there in Kungsträdgården.  
 ‘Maybe he buried in some way also his own youth there in Kungsträdgården.’

- c. *Men Lise Maria kan jag kanske lite hjälpa sen.* **OF**  
 But Lise Maria can I maybe bit help afterwards.  
 ‘But maybe I can help Lise Maria a little bit afterwards.’

- d. *Innan jag går vidare är det kanske bäst att* **AF**  
 Before I go further is it maybe best that  
*jag säger några ord om mig själv.*  
 I say some words about my self.  
 ‘Maybe it is good that I tell something about myself before I continue.’

The third clause type (henceforth clause type III) is a non-V2 clause which has *kanske* as its second constituent, that is, *kanske* occupies the position of the finite verb (135a). In these clauses *kanske* has the subordinating properties of a verb phrase because it is followed by subordinate word order. However, it is neither possible to paraphrase *kanske* as the verb phrase *det kan ske att*, as in (135b), nor to substitute *kanske* by synonymous adverbs, for example *sannolikt* ‘probably’, as in (135c).

- (135) a. *Han kanske inte kommer idag.*  
 He maybe not comes today.

<sup>82</sup> Some items, such as *Nåja* (a) and *Men* (c), are located in the so-called ‘Prefoundation’ because they are syntactically independent units, cf. Section 2.1.4 and Chapter 5 on the status of the discourse marker *I think*.

- b. \**Han det kan ske att inte kommer idag.*  
 \*He it may happen that not comes today.
- c. \**Han sannolikt inte kommer idag.*  
 \*He probably not comes today.

The third clause type includes non-V2 clauses that have *kanske* as its second constituent. There are three subtypes of clause type III, i.e. SK, OK and AK clauses. These non-V2 structures can have a subject, object or various types of adverbs/adverbials as their first constituent, as illustrated in (136) below.

- (136) a. *Vi kanske borde lägga Brita* **SK**  
 We maybe should lay Brita  
*lite mer bekvämt förresten.*  
 little more comfortable anyway.  
 ‘Anyway, maybe we should lay Brita a little more comfortable.’
- b. *-Vad jag tycker är humanitära skäl* **OK**  
 -What I think are humanitarian grounds  
*kanske inte min kollega håller med om,*  
 maybe not my colleague agrees with on,  
*vi är ju inga robotar.*  
 we are sure no robots.  
 ‘What I consider to be humanitarian grounds, maybe my colleague doesn’t consider to be humanitarian grounds, after all we are no robots.’
- c. *Nu när vi är tre man kanske vi skulle* **AK**  
 Now when we are three man maybe we could  
*kunna ta och palla upp bilen igen.*  
 can take and lift up car.the again.  
 ‘Now that we are three, maybe we can lift up the car again.’

All in all, there are different V2 and non V2 structures in which *kanske* may occur. The frequencies with which they occur, but also the different subtypes differ for the Mainland Scandinavian languages, as will be shown in the next section.

#### 4.4.3 Results

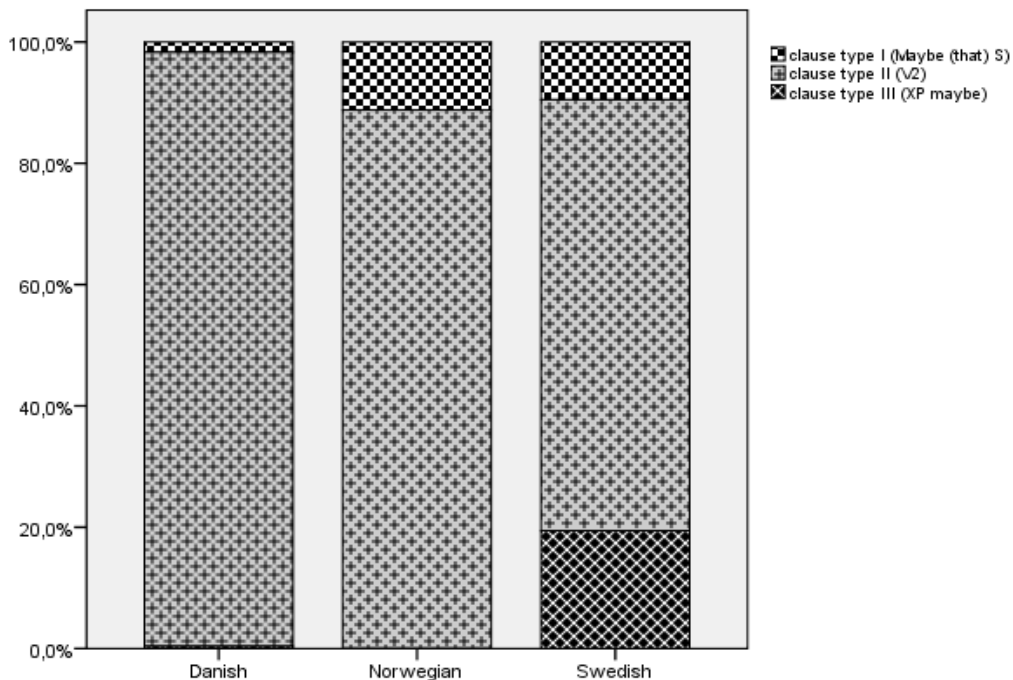
In this section the results of the replication studies on the syntactic distribution of Danish and Norwegian MAYBE are presented and contrasted with Andréasson’s findings for Swedish.

Swedish *kanske* occurs 5183 times in RII and 5913 times in P98. The random sample consists for 60.4 % of declarative main clauses; 71.1 % of these are V2 clauses. Norwegian *kanskje* occurs 196.504 times in *Norsk Aviskorpus*. As many as 78.4 % of the clauses in the random sample are declarative main clauses; 88.8 % of which are V2 clauses. Danish *måske* occurs 42.067 times in *KorpusDK*. About half of the clauses in the random sample are declarative main clauses, 52.7 %, and these consist almost solely of V2 clauses 98.3 %.

Clauses that do not adhere to the V2 principle are far less frequent than V2 clauses. Of the Swedish clauses, 28.9 % are non-V2 structures (clause type I (9.5%) and III (19.4 %)). Only 11.2 % of the Norwegian and 1.7 % of the Danish clauses are non-V2 structures (clause type I). The distributions of the three different clause types in the Mainland Scandinavian languages are shown in Figure 8.

A Chi square analysis of the data shows a significant relation ( $\chi^2=306.2$ ,  $df=4$ ,  $p<0.01$  and Cramer's  $V=0.290$ ) for language (Danish, Norwegian, Swedish) and clause type (I, II, III). That is, the distribution of the three different clause types is significantly different in Danish, Norwegian and Swedish, which indicates different degrees of grammaticalization in the Mainland Scandinavian languages. In all three languages, MAYBE occurs most frequently in V2 clauses (clause type II). Clause type I, which represents one of the older stages in the development of MAYBE, is marginally attested in all languages. Clause type III is only productive in Swedish. The Norwegian and Danish samples contained no instances of this structure. See Appendix 4 for an overview of the counts per clause type, as well as the different subtypes, in each language.

A Chi-square analysis of the data shows a significant relation ( $\chi^2=306.2$ ,  $df=4$ ,  $p<0.01$  and Cramer's  $V=0.290$ ) for language (Danish, Norwegian, Swedish) and clause type (I, II, III). That is, the distribution of the three different clause types is significantly different in Danish, Norwegian and Swedish, which indicates different degrees of grammaticalization in the Mainland Scandinavian languages. In all three languages, MAYBE occurs most frequently in V2 clauses (clause type II). Clause type I, which represents one of the older stages in the development of MAYBE, is marginally attested in all languages. Clause type III is only productive in Swedish. The Norwegian and Danish samples contained no instances of this structure. See Appendix 4 for an overview of the counts per clause type, as well as the different subtypes, in each language.



**Figure 8.** Bar chart of the distribution of clause type I, II and III for MAYBE.

#### 4.4.3.1 Clause type I

Clause type I, non-V2 clauses introduced by MAYBE, contains two subtypes. KaS with the complementizer THAT and KS clauses without complementizer. KS clauses are attested in Norwegian, Danish and Swedish, but this non-V2 structure is far less frequent than the standard V2 clauses. KS clauses form 9.5% of the Swedish, 11.3 % of the Norwegian and only 1.7 % of the Danish MAYBE clauses. Corpus examples of KS clauses are given in (137).

- (137) a. *Kanske*    *han*    *också*    *haft*    *en*    *arbetsam*    *dag.*                      Swedish  
           Maybe    he        also    had    a        hard-working   day.                      SK\_RII  
           ‘Maybe he also had a tough day.’
- b. *Kanskje*    *jeg*    *kan*    *omformulere*    *spørsmålet.*                      Norwegian  
           Maybe    I        can    reformulate    question.the.                      NAK  
           ‘Maybe I can reformulate the question.’

- c. *Måske*      *parterne*      *kunne blive*      *enige*      *om*      Danish  
 Maybe      parties.the      could become      agreed      on      KDK  
*en form for fælles-administration.*  
 a form for common.administration.  
 ‘Maybe the parties could come to an agreement on some  
 kind of joint administration.’

No occurrences of KaS clauses are attested in the Norwegian, Swedish and Danish random samples. Separate searches result in sporadic corpus examples, as shown in (138). The percentages<sup>83</sup> of KaS clauses are 0.1% in P98, 0.5% in BII, 0.02% in Norsk Aviskorpus and 0.0% in KorpusDK.

- (138) a. *Kanske att förbannelsen ännu inte är hävd.*      Swedish  
 Maybe that curse.the yet not is cancelled.      SK\_RII  
 ‘Maybe the curse is not yet cancelled.’
- b. *Kanskje at det ikke er så mye*      Norwegian  
 Maybe that it not is so much      NAK  
*som skjer i byen.*  
 that happens in town.the.  
 ‘Maybe there is not much going on in town.’
- c. *Måske at jeg kunne have afbrudt dem.*      Danish  
 Maybe that I could have interrupted them.      KDK  
 ‘Maybe I could have interrupted them.’

The coexistence of KS and KaS clauses can be viewed from two different angles. Either KS clauses are a reduced variant of KaS clauses (the complementizer THAT can be omitted which makes it a remnant of the original verb phrase IT CAN/MAY HAPPEN THAT) or KaS clauses are a variant of KS clauses (the complementizer THAT is optional between MAYBE and the subject).

This latter option presumes that KS clauses are older than KaS clauses (but see Wessén 1968; Andréasson 2002 for a description of the opposite situation). However, since THAT in general is an optional complementizer (Teleman et al. 1999:536f.), see example (139), this latter option is not implausible.

- (139) a. *Han sa*      *Svensson skulle komma.*  
 He said      Svensson would come.

<sup>83</sup> These percentages are based on the total number of occurrences of KaS clauses divided by the total number of occurrences of MAYBE in the entire corpus.

- b. *Han sa att Svensson skulle komma.*  
He said that Svensson would come.
- c. *Kanske han kommer.*  
Maybe he comes
- d. *Kanske att han kommer.*  
Maybe that he comes

The complementizer may have been (optionally) reintroduced at the time MAYBE was reanalyzed as a sentence adverb, which makes current KaS clauses a later development. In this scenario KaS clauses are a case of ‘insubordination’ (Evans 2007), a phenomenon that is widely attested in Germanic languages (Van de Velde & Van linden 2010). Some examples of insubordination are given in (140) below.

- (140) a. *Misschien dat iemand me kan helpen.* Dutch  
Maybe that someone me can help  
‘Perhaps someone is willing to help me.’
- b. *Schön dass du gekommen bist.* German  
Nice that you come is.  
‘How nice of you to come.’
- c. *Kjedelig at det gikk slik.* Norwegian  
Boring that it went so.  
‘Too bad it went like this.’

A more thorough analysis of diachronic data will clarify whether KaS clauses, as we find them today, are a remnant of the oldest stage in MAYBE’s development or a new development known as ‘insubordination’ or ‘semi-autonomous insubordination’, see Norde (2012) and Norde, Beijering & Rawoens (2012). This latter scenario is strengthened by the fact that more and more epistemic adverbs and other elements are possible in this construction.

#### 4.4.3.2 Clause type II

Clauses in which MAYBE is preceded by the subject and finite verb are very frequent. This comes as no surprise since Mainland Scandinavian languages are V2 languages and this clause type adheres to the V2 principle. SF clauses are most frequent in Norwegian. In Swedish and Danish, SF clauses come in second place. Examples of SF, KF, OF and AF clauses are given in (141).

- (141) a. *Man är kanske rädd att barnen* Swedish  
One is maybe scared that children.the SK\_P98

*inte ska lära sig svenska ordentligt.*  
 not shall learn themselves Swedish properly.  
 ‘Maybe they are afraid that the children do not acquire  
 the Swedish language properly.’

- b. *Jeg har kanskje gjort det motsatte, sier Gashi.* Norwegian  
 I have maybe done the opposite, says Gashi. NAK  
 ‘Maybe I have done the opposite, says Gashi.’

- c. *Han er måske lidt på den gamle side - over 60.* Danish  
 He is maybe little on the old side - over 60. KDK  
 ‘He is maybe a little bit too old - over 60 years of age.’

Like SF clauses, KF clauses are V2 clauses, see (142) below. In Swedish and Danish this structure is most frequent of all MAYBE clauses. For Norwegian this structure comes in second place.

- (142) a. ***Kanske*** *är det för att jag* Swedish  
 Maybe is it because that I SK\_P98

*är svart, jag vet inte.*  
 am black, I know not.  
 ‘Maybe it is because I am black, I do not know.’

- b. ***Kanskje*** *kommer mellom 5000* Norwegian  
 Maybe come between 5000 NAK

*og 10 000 tilskuere.*  
 and 10.000 spectators.  
 ‘Maybe there will be between 5000 and 10.000 people.’

- c. ***Måske*** *får du et tilbud fra Copenhagen* Danish  
 Maybe get you an offer from Copenhagen KDK

*Models, fnisede Marianne og så*  
 Models, chuckled Marianne and looked  
*lidt misundelig ud.*  
 little jealous out.  
 ‘Maybe you will get an offer from Copenhagen Models,  
 chuckled Marianne and she looked a bit jealous.’

OF clauses, i.e. clauses with a fronted object or predicate complement are far less frequent than KF and SF clauses in all three languages. Corpus examples are given in (143) below.



- (143) a. *Plågorna kan man kanske lindra något.* Swedish  
Pain can one maybe relieve little. SK\_RII  
‘One can maybe relieve the pain a little.’
- b. *Organene han fikk operert inn var kanskje* Norwegian  
Organs-the he got operated in were maybe NAK  
*de som fungerte best.*  
those that functioned best.  
‘The organs he was operated on were maybe those that functioned best.’
- c. *Dem kan vi med vores viden* Danish  
Them can we with our knowledge KDK  
*måske hjælpe til at kæmpe videre.*  
maybe help with to fight further.  
‘With our knowledge we can maybe help them to keep on fighting.’

Also AF clauses i.e. clauses with fronted adverbials are less frequent than KF and SF clauses. Corpus examples are given in (144) below.

- (144) a. *Nu blir jag kanske avstängd nästa* Swedish  
Now become I maybe suspended next SK\_P98  
*match, säger Matti*  
match, says Matti.  
‘Maybe I will now be suspended the next match, says Matti’
- b. *Onsdag får jeg kanskje begynne* Norwegian  
Wednesday get I maybe begin NAK  
*å trene lett igjen.*  
to train easy again.  
‘On Wednesday maybe I can start training lightly again.’
- c. *Før eller siden får vi måske at vide,* Danish  
Sooner or later get we maybe to know, KDK  
*hvad der virkelig skete i computerrummet*  
what there really happened in computer room-the  
*på USS Vincennes.*  
at USS Vincennes.  
‘Sooner or later we might get to know what really happened  
in the computer room at USS Vincennes.’

All in all, V2 clauses with either the subject or MAYBE as the first constituent are much more frequent than V2 clauses with a preposed object or adverbial in the foundation.

#### 4.4.3.3 Clause type III

A striking difference between Swedish on the one hand and Norwegian and Danish on the other hand, is that OK, AK, and SK clauses are all part of the Swedish random sample whereas they are not attested in the Norwegian and Danish random samples. These clauses are also not mentioned in standard grammars of Norwegian and Danish, *Norsk Referanse Grammatikk* (Faarlund et al. 1997) and *Dansk Grammatik* (Zola-Christensen & Christensen 2005) respectively. A separate search for these subtypes results in a few Norwegian examples (only SK and AK clauses), but no Danish examples. Hence, unattested clause types are not necessarily non-existent but they may occur very rarely. Examples of SK, OK and AK clauses are given in (145) below.

- (145) a. *Hon kanske inte alls känner* Swedish  
 She maybe not all recognize SK\_RII  
*igen mej, säger hon tyst.*  
 again me, says she softly.  
 ‘Maybe she does not recognize me at all, she says softly.’
- b. *Han kanskje skadet seg da han* Norwegian  
 He maybe hurt himself when he NAK  
*falt, var tiltaltes forklaring på hvordan* Kildebo  
 fell was suspect.the’s statement of how Kildebo  
*havnet bevisstløs og sterkt forslått i veibanen.*  
 ended unconscious and strongly beaten in road.the.  
 ‘Maybe he hurt himself when he fell, was the suspect’s statement of  
 how Kildebo ended up unconscious and strongly beaten up on the road.’
- c. *Och helt ogrundade kanske* Swedish  
 And completely unfounded maybe SK\_P98  
*inte de misstankarna är.*  
 not these suspicions are.  
 ‘These suspicions are maybe not completely unfounded.’
- d. *När någon av våra idéer bliver verklighet* Swedish  
 When one of our ideas become reality SK\_P98  
*kanske inte vi går på skolan längre.*  
 maybe not we go on school.the longer.  
 ‘When one of our ideas come true, we might not be  
 going to school any longer.’

- e. *Da kanskje du blir litt mer populær.* Norwegian  
 Then maybe you become bit more popular. NAK  
 ‘Then you might become a little more popular.’

The status of AK, OK and SK clauses is unclear because these clauses seem to have emerged suddenly in literary language use. Especially SK clauses are very frequent in the work of the Swedish writer August Strindberg (1849-1912), whereas they were very infrequent in the period just before and after his work was published (Andréasson 2002:32). Nowadays, this clause type is well established in Swedish, 9.5% of the declarative main clauses, but extremely rare in Norwegian and possibly non-existent Danish. More research on the status of clause type III is needed to find out whether its development is a ‘natural’ development or an instance of coinage (invention of a new structure).

#### 4.4.4 Mainland Scandinavian MAYBE: a case of lexicalization, grammaticalization or pragmaticalization?

As we have seen in the previous sections, the development of Mainland Scandinavian MAYBE is a change from major (VP) to minor category (ADV), and a successive shift within the adverbial domain into speech-act adverb. In what follows, I will discuss and analyze the development of Mainland Scandinavian MAYBE in relation to the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization and (inter)subjectification.

First the essential mechanisms in lexicalization, grammaticalization and pragmaticalization will be discussed in subsection 4.4.4.1, then the accompanying primitive changes will be examined in subsection 4.4.4.2, the side effects that may identify potential instances of a certain X-ization are elaborated on in subsection 4.4.4.3, and finally, in 4.4.4.4 I will comment on the linguistic status (i.e. lexical, grammatical or communicative) and degrees of lexicalization, grammaticalization or pragmaticalization for Mainland Scandinavian MAYBE.

##### 4.4.4.1 Mechanisms in the development of MAYBE

As defined in the present study, lexicalization, grammaticalization and pragmaticalization are essentially conceived of as the result of formal reanalysis and semantic reinterpretation. With regard to MAYBE, the scheme looks as in Table 30. The cells that apply to MAYBE are shaded.

The development of the epistemic adverb MAYBE essentially involves the reanalysis from verb phrase to sentence adverb. MAYBE may also function as speech-act adverb, in which case there has been a successive intra-categorical reanalysis from epistemic to speech-act adverb. It also involves a constituent internal reanalysis (i.e. univerbation) because a syntagm is reduced to a simple lexeme, see (146) below.

- |       |                     |           |             |         |
|-------|---------------------|-----------|-------------|---------|
| (146) | (det) kan ske (att) | → kan ske | → kanske    | Swedish |
|       | VP                  | VP/sADV   | sADV/s-aADV |         |

Semantically, there are contiguous, metonymic relations between the more concrete ‘may happen’, the more general ‘may be’, and the speech-act use ‘may be?’ in interrogative clauses. An in depth historical survey may reveal the small steps that eventually led to these changes in meaning as well as the bridging-contexts for these semantic changes.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	±
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-
<b>reinterpretation</b>					
<b>-metaphor/metonymy</b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

**Table 30.** Mechanisms in the development of MAYBE.

#### 4.4.4.2 Primitive changes in the development of MAYBE

Formal reanalysis and semantic reinterpretation are accompanied by a subset of correlated primitive changes that may, but need not be involved in lexicalization, grammaticalization or pragmaticalization. The primitive changes that apply to MAYBE are represented by the shaded cells in Table 31.

In the development of MAYBE, the levels of phonology/phonetics and morphology are both affected. The development of MAYBE is a clear case of univerbation that is accompanied by phonetic reduction and which results in a reduced form (i.e. syntagm > lexeme).

In Norwegian and Swedish, but not in Danish, there has been a shift in stress. Norwegian *kanskje* and Swedish *kanske* have stress on *kan* but no stress on *skje/ske*. *Måske* takes stress on *ske*. In case of univerbation the original prosodic structure is lost. The original verb phrase accentuation (*kan'ske*) changed into compound accentuation (*'kanske*). At the beginning of the 19th century *kanske* could be pronounced in both ways. In the middle of the 19th century it is only mentioned with compound accentuation in dictionaries but in SAOB it is written that it is possible to pronounce *kanske* with verb phrase accentuation, for example in poetry (Andréasson

2002:37). The inflectional properties and variability of a verb phrase got lost when the unverbated adverb *kanske* arose.

The reanalysis from verb phrase to epistemic adverb results in increased syntactic variability and autonomy. Adverbs in general are more flexible and less tightly integrated into the syntactic structure than VP's. At first, the adverbialization of MAYBE results in more syntactic variation. That is, MAYBE may occur in positions in which it is followed by subordinate word order (=non V2 structures) and in positions for sentence adverbs (=V2 structures). After a while of co-existence, subordinating MAYBE in non-V2 clauses is gradually being replaced by adverbial MAYBE in V2 clauses, see Figure 8.

The meaning of MAYBE is bleached in the sense that it has become more general. That is, instead of stating that something 'may happen' it came to denote epistemic possibility 'maybe'. The semantic compositionality of MAYBE is transparent in the sense that its subparts are still recognizable. However, the meaning of *sk(j)e* 'happen' got lost and replaced by the more general 'be'. Moreover, a subjective component is added to its meaning as there is a development from an objective, declarative predicate that states that something may happen (It may happen that X) to an adverb that expresses the speaker's subjective evaluation of the likelihood of the predicate (Maybe X). As regards the epistemic and speech-act adverb there has been a shift from syntactic to speaking subject, which is not yet present in the declarative predicate 'it may happen that.' The role of (inter)subjectification will be discussed in more detail in subsection 4.4.4.4.

<b>ii. Primitive changes</b>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>-phonology/phonetics</b>				
° loss of phonological/phonetic substance	(+)	(+)	(+)	(+)
<b>-morphology</b>				
° loss of morphological compositionality	+	(+)	(+)	(+)
° loss of morphosyntactic properties	-	+	(+)	(+)
<b>-syntax</b>				
° loss of syntactic variability	-	+	+	-
° loss of syntactic autonomy	-	+	+	-
<b>-semantics</b>				
° loss of semantic substance	-	+	+	+
° loss of semantic compositionality	+	(+)	(+)	(+)
<b>-discourse/pragmatics</b>				
° subjectification	(+)	(+)	(+)	+
° intersubjectification	(+)	(+)	(+)	+

**Table 31.** Primitive changes in the development of MAYBE.

#### 4.4.4.3 Side effects in the development of MAYBE

The side effects of formal reanalysis and semantic reinterpretation and their accompanying primitive changes can be used as a diagnostic to identify potential cases of lexicalization, grammaticalization or pragmaticalization. As regards MAYBE, the side effects that are noticeable in its development are represented by the shaded cells in Table 32 below.

<u>iii. Side effects of change</u>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
-paradigmaticization	-	+	+	(+)
-obligatorification	-	(+)	(+)	-
-condensation	-	+	+	-
-layering/divergence/specialization/persistence	+	+	+	+
-productivity	-	+	+	+
-frequency	-	+	+	+
-typological generality	-	+	(+)	(+)

**Table 32.** Side effects in the development of MAYBE.

The side effects *paradigmaticization*, *productivity* and *frequency* are closely related to each other. Through the course of its development, MAYBE enters the adverbial paradigm, more specifically the class of epistemic adverbs. As such its adverbial positions and functions are extended. Besides being an epistemic sentence adverb, MAYBE can also be a modifying adverb or a speech-act adverb. Nowadays MAYBE is one of the most frequent epistemic adverbs in Mainland Scandinavian, cf. Table 16. The verb phrase (IT)CAN/MAY HAPPEN (THAT) is far less frequent these days.

*Obligatorification* does not apply to MAYBE. There are various ways to express epistemic dimensions, adverbs are just one of them, as such MAYBE has not become an obligatory part of grammar. Likewise *condensation* does not occur because the structural scope of MAYBE is not reduced, nor does it become (more) dependent upon other constituents in the clause, rather the opposite. That is, sentence adverbs are integrated into syntactic structures, but flexible with regard to their positions. The univerbation of a modal form ‘may/can’ and main verb ‘happen/be’ into an epistemic adverb is not unique to Mainland Scandinavian MAYBE, but a widely attested tendency in other European languages too, (Section 4.2.1).

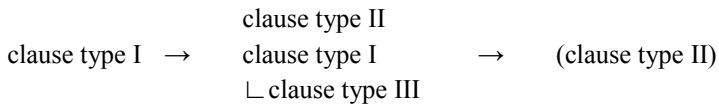
*Layering* provides a good description of the coexistence of adverbial and subordinating MAYBE. The older MAYBE with subordinating properties still exists along with the newer MAYBE that functions as pure sentence adverb. The new layer (clause type II) gradually replaces the older layer (clause type I), as illustrated in (147).

(147) clause type I      →      clause type I ~ clause type II      (→      clause type II)

Completed adverbialization for MAYBE would mean that it only occurs as pure sentence adverb in V2 clauses, which is not yet the case but Danish is getting close, see Figure 8. It is the tendency that non-V2 structures are gradually being replaced by V2 structures, but it should be noted that language change is by no means deterministic. Therefore clause type II as the final stage of the development of MAYBE is put into brackets. In V2 structures MAYBE is most adverbialized because in these structures it has lost the subordinating properties of the original verb phrase and functions either as an epistemic marker of possibility or speech-act adverb.

Clause type III is a common structure in Swedish but extremely rare in Danish and Norwegian. This may be instance of divergence or split, but it is only relevant in the development of MAYBE if the new variant of subordinating MAYBE (clause type III) is not an instance of coinage. Recall that this clause type first emerges in literary and poetic texts, and is especially very frequent in the work of August Strindberg who might have introduced this structure to a large audience. If the rise of this clause type is a ‘natural’ development then it forms a new layer out of the original verb phrase or K(a)S clauses, as shown in (148). Note that this clause type is not an intermediate stage because both subordinating and adverbial MAYBE already existed before this clause type came into being.

(148)



When MAYBE was not yet an univerbated adverb, other modal forms could be used as alternating forms for *kan*, for example *må* ‘may’ or *tör* ‘must’. The main verb *ske* ‘happen’ could alternate with *hända* ‘happen’ and other words like for example the modal particle *väl* ‘well’ could be inserted between the modal form and the main verb. Inflection for tense was also possible in the older stages of development. Out of these options the synonymous adverbs *måhända*, *törhända* and *måhända* were formed. *Kanske* is by far the most frequent epistemic adverb of the MAYBE-type in Swedish (5913 hits in P98 and 5183 hits in RII), *kanhända* and *måhända* occur less frequently (respectively 23 and 92 hits in P98, 52 and 85 hits in RII) and *törhända* is archaic but may be used in literary texts (1 hit in P98 and 35 hits in RII). Danish and Norwegian counterparts of *måhända* and *törhända* are not found in *KorpusDK* and *Norsk Aviskorpus*. The counterpart of *kanhända* in Danish and Norwegian can be written as one or two words, both forms are correct. Univerbated *kanhænde* occurs 4 times in *KorpusDK* and *kanhende* occurs 55 times in *Norsk Aviskorpus*. The non-univerbated form *kan hænde* occurs 57 times in *KorpusDK* and *kan hende* 5198 times in *Norsk Aviskorpus*. This is an instance of *specialization*, i.e. the phenomenon that many similar and simultaneously existing expressions are reduced to one major expression, as MAYBE is now the most frequent adverb to express epistemic possibility.

As regards *persistence*, it has been mentioned several times that MAYBE in non-V2 clauses has retained subordinating properties of a verb phrase. This explains the violation of the

V2 principle in clause type I. For clause type III there is no definite answer to this question. The V2 violation may be due to its verbal origin or it could be an instance of coinage. In case the rise of clause type III is a ‘natural’ development, a plausible developmental path is proposed by Andréasson (2002:31).

She shows how this clause type could have developed by reconstructing an older stage of AK clauses in (149). The example in (149a) is the original example and (149b) is the reconstructed example. In AK clauses, it is possible to paraphrase *kanske* (*kan skie* in (149a)) as verb phrase (*kan det skie att* (149b)). Through the course of time, the possibility to paraphrase *kanske* into a verb phrase was lost, but *kanske* could still occur in this non-canonical position for sentence adverbs. The other subtypes of clause type III (SK and OK clauses) could have developed by analogy with AK clauses.

- (149) a.    *och*    *om*    *jag*    *det*    *än*                    *woro*,  
               and    if    I        it        anyway        were,  
               *så*    *kan*    *skie*                    *det*    *hulpo*            *intet*.  
               than   may   happen            it        helped            nothing.  
               ‘and if I were that, so maybe it did not help.’
- b.    *och*    *om*    *jag*    *det*    *än*                    *woro*,  
               and    if    I        it        anyway        were,  
               *så*    *kan*    *det*    *skie*                    *att*    *det*    *hulpo*            *intet*.  
               so    may   it        happen            that   it        helped            nothing.  
               ‘and if I were that, so maybe that it did not help.’

Hence, for clause type III there is no clear-cut answer as to why it violates the V2 principle. Different ideas have been proposed in order to clarify the V2 violation in clause type III, but only a detailed historical survey may reveal which scenario is most plausible.

#### 4.4.4.4 Linguistic status and degrees of X-ization for MAYBE

As defined in the present study, lexicalization results in a linguistic item belonging to a major category, with referential meaning, primary status, and which may convey the main point of linguistic message, grammaticalization leads to a linguistic item belonging to a minor category, with relational meaning, secondary status, and which regulates grammatical structures and relations, and pragmaticalization results in a linguistic item with communicative meaning, extra-propositional status, and which organizes discourse structure.

With regard to MAYBE, the over result of the changes is an adverb that denotes epistemic possibility and related speech-act functions. The patterns that arise from the Table 30-Table 32 show that its development has characteristics of lexicalization, grammaticalization and pragmaticalization. As regards the general development, it has most properties in common with both primary and secondary grammaticalization. With respect to semantic and morphological



compositionality there is overlap with lexicalization I (fusion), and with regard to syntax and discourse/pragmatics it sides with pragmaticalization

The development of the epistemic adverb is an instance of grammaticalization accompanied by subjectification at the textual level. In case of the speech-act adverb there has been an intra-categorical shift in the adverbial domain which is accompanied by intersubjectification. Table 33 summarizes the types of subjectification in MAYBE.

<b>(Inter)Subjectification</b>	<b>MAYBE</b>
<b>I. subjectification</b> [speaker perspective, attitude and judgment]	+
- <b>textual level</b> [meta-linguistic meanings] epistemic meaning	
<b>II. intersubjectification</b> [interaction with interlocutor]	+
- <b>textual level</b> speech-act meanings	

**Table 33.** Types of subjectification in the development of MAYBE.

In the overall development, Danish has progressed most in that *måske* is primarily an epistemic/speech-act adverb. This is also reflected by the fact *måske* almost solely occurs in V2 clauses. Norwegian, and especially Swedish, allow considerably more non-V2 clauses with respect the MAYBE.

#### 4.5 Overall summary, discussion and conclusions

The results of the case study on Mainland Scandinavian MON highlight considerable differences, both formally and semantically, with respect to its status and development. The interrogative adverb MON expresses various dimensions of doubt from the point of view of the speaker and may occur in different syntactic constructions.

In Danish, *mon* often co-occurs with a *wh*-word or the negation marker *ikke* ‘not’. Overt conjunctions are not attested in the sample. *Mon* may also function as an epistemic sentence adverb. In Norwegian the interrogative adverb *mon* mainly occurs together with the cognitive verb *tro* ‘think, believe’ (Chapter 5), which is optionally followed by an overt conjunction. Swedish *månne* functions as an epistemic/speech-act sentence adverb in interrogative clauses. Because of the reanalysis from interrogative adverb to sentence adverb, V2 word order is possible in clauses with *månne* as the first constituent. This is not possible for Danish and Norwegian as *mon* is primarily an interrogative adverb, occasionally it occurs as a sentence adverb in Danish.

In explaining the complex syntactic behavior of MON it has been shown to be important to take into account diachronic facts of its development. In non-V2 structures, MON retained traces of the auxiliary form MONNE. As an interrogative adverb it has subordinating properties because a conjunction like IF or a *wh*-word is (covertly present). In V2 structures MON has lost

the subordinating properties of the question particle and functions as an epistemic sentence adverb or speech-act adverb.

The results of the case study on Mainland Scandinavian MAYBE highlight substantial differences with respect to its syntactic distribution. The distribution of the three different clause types is significantly different in the Mainland Scandinavian languages, see Figure 8. In Swedish there are more syntactic positions available than in Danish and Norwegian, one of which is the place of the finite verb. The reason why Swedish *kanske* can occupy the position of the finite is most likely due to the fact that it belongs to the class of clausal words and thus has preserved verbal properties. For Norwegian and Danish it is unclear if clause type III ever existed to the same extent as in Swedish. Closer scrutiny of the status of SK, OK and AK clauses will make clear whether this is a case of divergence (split) or coinage, i.e. invention and introduction of a new clause type in the language. All in all, MAYBE functions more and more as pure sentence adverb and occurs most frequently in V2 clauses in the Mainland Scandinavian languages. MAYBE is losing its subordinating properties and non-V2 clauses are gradually being replaced by V2 clauses in which MAYBE functions as pure sentence adverb.

Both MON and MAYBE display remarkable syntactic behaviour. They may violate the V2-principle, which can be explained by taking into account their historical development and verbal origin. The differences and similarities between the development of MON and MAYBE are summarized in Table 34. White cells denote that a certain mechanism/primitive change/side effect is not applicable to both MON and MAYBE, the middle or 50 % grey cells denote commonalities between MON and MAYBE, the light or 25% grey cells are applicable to MON only, and the dark or 75 % grey cells apply to MAYBE only.

Both MON and MAYBE have epistemic and speech-act meanings but their development differs in many respects. The main difference is that MON is a shift from minor to minor category and MAYBE from a major to minor category. The development of MAYBE also involves a constituent internal reanalysis because it is a univerbation of a modal form and a main verb, as shown Table 34 below.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	±
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<i>reinterpretation</i>					
<b>-metaphor/metonymy</b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

**Table 34.** Mechanisms in the development of MON and MAYBE.

The primitive changes that accompany the formal reanalysis and semantic reinterpretation coincide largely for MAYBE and MON, see Table 35 They differ only with respect to loss of phonetic/phonological substance and morphological compositionality, which is due to the fact that MON is a monomorphemic item and MAYBE a compositional form.

<b>ii. Primitive changes</b>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>-phonology/phonetics</b>				
° loss of phonological/phonetic substance	(+)	(+)	(+)	(+)
<b>-morphology</b>				
° loss of morphological compositionality	+	(+)	(+)	(+)
° loss of morphosyntactic properties	-	+	(+)	(+)
<b>-syntax</b>				
° loss of syntactic variability	-	+	+	-
° loss of syntactic autonomy	-	+	+	-
<b>-semantics</b>				
° loss of semantic substance	-	+	+	+
° loss of semantic compositionality	+	(+)	(+)	(+)
<b>-discourse/pragmatics</b>				
° subjectification	(+)	(+)	(+)	+
° intersubjectification	(+)	(+)	(+)	+

**Table 35.** Primitive changes in the development of MON and MAYBE.

The peculiar status of MON is highlighted by the fact that it does not enter a paradigm (i.e. there is no class of items serving the same functions), it does not become more frequent, and its development is unique to the Mainland Scandinavian languages. This is the opposite for MAYBE which becomes the most prominent member of the paradigm of epistemic adverbs, and as such it is a highly frequent item. The side effect for MON and MAYBE are given in Table 36.

<b>iii. Side effects of change</b>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
-paradigmaticization	-	+	+	(+)
-obligatorification	-	(+)	(+)	-
-condensation	-	+	+	-
-layering/divergence/specialization/persistence	-	+	+	+
-productivity	-	+	+	+
-frequency	-	+	+	+
-typological generality	-	+	(+)	(+)

**Table 36.** Side effects in the development of MON and MAYBE.

MON and MAYBE have many side effects, e.g. layering, divergence, specialization, persistence, in common that are simply due to the fact that any categorical reanalysis results in an old en new form that remain related in some way.

## Chapter 5

### 5. Case study III: Mental state predicates TROR

#### 5.1 Introduction

This chapter presents a comparative synchronic corpus investigation of the formal and semantic properties of the mental state predicate ‘I think’ in the Mainland Scandinavian languages. These are Danish *jeg tror*, Norwegian *jeg tror* and Swedish *jag tror* and are referred to as Mainland Scandinavian TROR whenever all three of them are discussed together. The focus of this chapter is the development of the discourse marker TROR in relation to the lexicalization-grammaticalization-pragmaticalization interface and (inter)subjectification.

The status of discourse markers is a much disputed issue in grammaticalization studies as their development has been termed lexicalization, grammaticalization and pragmaticalization (see Section 2.2.4). Parenthetical phrases such as TROR have been subsumed under lexicalization because they can be seen as formulaic tokens. They have been included in (broad definitions of) grammaticalization because they generally conform to some well-known grammaticalization criteria (while at the same time violating other basic principles of grammaticalization). And, they have been assigned a special status that is different from both lexicalization and grammaticalization, i.e. pragmaticalization, because discourse markers are linguistic items that operate at the discourse level and do not belong to the core grammar. That is, they mark discourse structure, not grammatical structure.

The aim of this chapter is to determine to what extent Mainland Scandinavian TROR, like English *I think* (Aijmer 1997:1), shows signs of becoming a discourse marker, and whether this development is a case of lexicalization, grammaticalization or pragmaticalization. The following research questions, as stated in (150), will be addressed.

- (150) What are the differences and similarities for the Mainland Scandinavian languages with respect to the etymology, development and syntactic and semantic properties of the mental state predicate TROR? Is its development a case of lexicalization, grammaticalization or pragmaticalization? And more specifically, do the Mainland Scandinavian languages differ in the degree of grammaticalization, lexicalization and/or pragmaticalization for TROR?

The status of Mainland Scandinavian TROR will be examined on the basis of the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization as shown in Table 10 in Chapter 2. The role of (inter)subjectification in the rise of TROR will also be discussed.

This chapter is organized as follows: Section 5.2 discusses discourse markers in general and mental state predicates of the type *I think* in more detail. In Section 5.3 the sources and method used in this case study are discussed. In Section 5.4, the results of a comparative corpus investigation of the mental state predicate TROR in Danish, Norwegian and Swedish are

presented. In Section 5.5, the lexicalization-grammaticalization-pragmaticalization interface and the role of (inter)subjectification will be discussed with respect to the development of Mainland Scandinavian TROR. Finally, Section 5.6 contains a summary, discussion and the conclusions of this case study.

## 5.2 Discourse markers

Discourse markers have been studied from different perspectives and within different approaches. Consequently, there is no consensus on a definition, terminology and classification of discourse markers. The class of discourse markers (also termed pragmatic markers, pragmatic particles or discourse particles) comprises a wide variety of linguistic items and expressions that belong to various types of word classes. According to Aijmer & Simon-Vandenberg (2009) the class of discourse markers includes, amongst others, connectives, modal particles, pragmatic uses of modal adverbs, routines (*how are you*), disjuncts (*frankly, fortunately*), pragmatic uses of conjunctions (*and, but*), and interjections. Hence, both formally and semantically, discourse markers do not form a coherent class of linguistic items.

General properties of discourse markers, which may single them out as a separate category, have been identified in the literature (e.g. Schifffrin 1987; Brinton 1996). Discourse markers are frequently used forms that primarily occur in spoken language. Because they are associated with informality, they are often stylistically stigmatized or undesired in written texts. Being primarily part of the spoken language, discourse markers are transient, relatively short-lived forms that are constantly subject to renewal.

Discourse markers appear to be empty or semantically reduced expressions that nevertheless serve a multitude of discourse functions. They tend to be syntactically free and optional elements, whose word class, distribution and meaning(s) are far from clear. They are marginal forms that are hard to translate and difficult to place in traditional word classes.

Most discourse markers can be analyzed as a subtype of (speech-act) adverbials. As discussed previously in Chapter 4, ‘adverb’ is a notoriously difficult category which consists of various subsets of grammatical, lexical and communicative variants. Discourse markers are not grammatically obligatory, but pragmatically required (cf. Diewald’s (2011) ‘communicative obligatoriness’). Omission of discourse markers does not render ungrammatical or unintelligible structures, but the language use appears to be unnatural or incomplete when (subtle) pragmatic dimensions are missing.

Discourse markers are not integrated into the syntactic structure of the clause, but rather loosely attached to it as meta-linguistic comments. Discourse markers are difficult to specify lexically, or in terms of grammatical functions. They operate at the interpersonal level, (see Section 2.1.1.4 on (inter)subjectification), and are therefore not part of the propositional content of the clause. Discourse markers relate to units larger than the sentence. Because of this, they are best analyzed with reference to discourse organization, the communicative context, politeness and other social and cultural factors. They are structural markers signaling speaker attitudes and discourse relations.

In the remainder of this chapter I will only be concerned with a particular subtype of discourse markers, namely ‘mental state predicates’ (Nuyts 2001) of the *I think*-type. These are also known as parenthetical verbs (Urmson 1952), first-person epistemic parentheticals (Brinton 1996), complement taking predicates (CTP, Boye & Harder 2007), or complement taking mental predicates (CTMP, van Bogaert 2010).

### 5.2.1 Mental state predicates of the *I think*-type

The set of mental state predicates is a relatively open class. It includes, inter alia, forms like *I think*, *I believe*, *I guess*, *I doubt*, *I suppose* and *I know*. The phrase *I think* is the most frequent and prototypical member of this category. Mental state predicates are made up of cognitive verbs expressing supposition, inference, expectation or belief and they generally occur with a first-person pronoun in simple present tense.

As pointed out by Nuyts (2002:107) mental state predicates are hard to deal with because of their complex semantic structure and intricate syntactic behavior. This section discusses the formal properties of *I think* in subsection 5.2.1.1, the semantic properties of *I think* are examined in subsection 5.2.1.2, and subsection 5.2.1.3 elaborates on proposed developmental paths for *I think* into a discourse marker.

#### 5.2.1.1 Formal properties of *I think*

The mental state predicate *I think* is a frequently occurring phrase of the general form ‘cognitive verb in simple present tense + first person pronoun’, which may, but need not be, omitted from an utterance. It may occur outside the syntactic clause, in which case it contributes little to the propositional meaning of the clause. For English *I think*, Aijmer (1997:1) observes that this phrase “seemed to function in the same way as modal particles in languages like German or Swedish [...] and developed into a discourse marker or modal particle which is syntactically a speech-act adverbial.” However, since *I think* is flexible with regard to tense, aspect, modality, negation and questioning it is not a pure pragmatic expression like *you know* or *you see*, which cannot be subject to these operations (ibid:6).

There are basically two syntactic patterns for mental state predicates (Nuyts 2001; Brinton 1996:228): a complementing pattern in which *I think* is positioned at the beginning of a clause, as in (151a), and a parenthetical pattern in which *I think* occurs in clause-medial or -final position, as in (151b).<sup>84</sup> In its complementing form, *I think* introduces a complement clause which may optionally contain the complementizer *that*. As a parenthetical, *I think* is best analyzed as an epistemic adverb expressing speaker attitude.

(151) a. ***I think*** (*that*) *the book was written by John.*

b. *The book, (***I think***), was written by John, (***I think***).*

<sup>84</sup> The status of *I think*, its complement and the entire construction has been subject to discussion in the literature (cf. van Bogaert 2010; Boye & Harder 2007). The phrase *I think* does not have the status of a clause, it only does in combination with its complement. The status of the main and complement clause is a point of discussion.

In the Germanic languages, with the exception of English, parenthetical instances of *I think* are characterized by inversion of the subject and finite verb. Parentheticals lack grammatical integration in the clause. That is, they are used as insertions or afterthoughts. The optionality of the phrase *I think* is probably best explained by its function, that is, whether it is complementing as in (151a), or modifying the content of the clause as in (151b).

### 5.2.1.2 Semantic properties of *I think*

Semantically, *I think* has a non-qualificational (=lexical or literal) meaning of being in mental state X or performing mental state X, as in (152a), and a qualificational (meta-linguistic or interpersonal) meaning that expresses the source of knowledge and/or a degree of likelihood, as in (152b), (cf. Nuyts 2001). The qualificational meaning of *I think*, as in (152b), conveys the speaker's epistemic/evidential stance rather than describing a concrete mental process.

- (152) a. *I am thinking.*
- b. *I think it happened this way.*

Mental state predicates are mixed epistemic-evidential forms (Nuyts 2001; Brinton 1996). In some cases the epistemic meaning may prevail, in other cases the evidential dimension may be more prominent. As observed by Nuyts, these notions are inseparable because qualificational *I think* is always ambiguous between the mental process of thinking/believing and the expression of belief.

The qualificational and non-qualificational meanings of *I think* can be distinguished in terms of transitivity. The non-qualificational variants are lexical verbs in a main clause construction. Moreover, they are often accompanied by prepositions, as in *think about/of* or *believe in*. The qualificational meanings occur either as a complement taking predicates (CTP) or as a parenthetical (DM). Qualificational mental state predicates can be omitted from the utterance without changing the meaning of the main message ([...] it happened this way). They may also be substituted by adverbial equivalents (e.g. Perhaps it happened this way). These adverbial equivalents lack the 'personal responsibility' that is typical of mental state predicates, as the use of mental state predicates "indicates that the speaker assumes strictly personal responsibility for the information provided" (Nuyts 2001:122). To reinforce this, subjectivity markers like 'personally' may be added to underline that the speaker is not voicing general accepted truths, i.e. there may be an opposition in opinion between the speaker and hearer.

Mental state predicates are commonly used for describing epistemic evaluations. They can be used performatively, descriptively or ascriptively (Nuyts 2001). When used performatively the mental state predicate expresses or reports on one's own thoughts and beliefs at the moment of speaking (*I think*). When used descriptively or ascriptively (*s/he thinks*) the mental state predicate expresses or reports on someone else's thoughts and beliefs. Both descriptive and ascriptive uses are instances of reported speech. The difference is that in case of



descriptive use the person reported on has indeed uttered ‘I think X’ in case of ascriptive use it is unlikely that the person reported on has really uttered ‘I think X’. In general, descriptive uses are indeterminate with respect to whether they are reporting on or ascribing thoughts to someone else.

Note that descriptive uses of mental state predicates also involve speaker judgments in the sense that the speaker evaluates thoughts of someone else. By using descriptive mental state predicates, the speaker may express contrast of opinion (*s/he thinks X, but I think Y*) or signal inaccurate assertions (*s/he thinks X; I know that X is not true*). Descriptive uses of mental state predicates tend to convey skepticism on behalf of the speaker with regard to the opinion or thoughts of someone else (Nuyts 2001). Descriptive uses generally occur in the complementing pattern.

In addition to expressing various qualificational and non-qualificational meanings, *I think* may be used to serve various discourse functions. It is both a speaker- and hearer-oriented expression. Nuyts (2001:162-7) identifies different discourse strategies of performative *think*. The mental state predicate *I think* may, inter alia, be used as a hedging device and/or face-saving device. When used this way, “the predicate weakens or mitigates the force of the claim or the reaction, in such a way that it does not endanger the conversation and leaves room for intervention by the interaction partner” (Nuyts 2001:165). A similar observation has been made by Brinton (1996:239), who describes two interpersonal uses of first person epistemic parentheticals. The first one is the use of *I think* as an ‘intimacy signal’ that calls upon the hearer to participate in the discourse. The second one relates to *I think* as a negative politeness marker, which allows the hearer freedom to respond in different ways.

All in all, the expression *I think* displays a variety of different meanings, functions and syntactic structures, which are all interrelated to one another. In section 5.2.1.3, some ideas from the literature about the development of *I think* will be discussed, in order to get an idea of how this formal, semantic and functional variation came into being.

### 5.2.1.3 Development of *I think*

In Brinton’s (1996:239) study on pragmatic markers in English, it is found that mental state predicates, as well as epistemic adverbs, seem to be of relatively recent origin. This is concluded on the basis of the observation that Old English, in general, is deficient in markers of epistemic modality (Goossens 1982). Different scenarios for the development of mental state predicates, especially the parenthetical variant, have been proposed in the literature (e.g. Thompson & Mulac 1991; Brinton 1996).

According to Brinton (1996:252), the source constructions of parentheticals are numerous and diverse, as presented in (153). While acknowledging that the link between syntactic and semantic change is hard to establish, she suggests that in Stage I, the cognitive act (*that I think*) and the content of the act (*they are poisonous*) have an equal status. In Stage II, the phrase *that I think* denotes a mode of knowing rather than a cognitive act and its role has become secondary with respect to the matrix clause. In Stage III, *I think* is appositionally rather than anaphorically

connected to the matrix clause which results in the loss of anaphoric connectives, which are present in Stage II (e.g. *that, as, so*). Instead of indicating the source of knowledge for a given proposition, *I think* now predominantly expresses the degree of (un)certainly of the proposition in question.

The structures exemplified in Stage IV show that syntactic and semantic shifts have occurred with respect to the development of *I think*. Note that *I think* in clause-initial position is ambiguous between a ‘parenthetical’ and a ‘non-parenthetical matrix clause with ‘that-deletion’ (*I think they are poisonous*). Brinton (1996:253) adds that from Stage II onwards it is possible “to express the cognitive act only by means of the regular main verb-complement structure (*I think [that] they are poisonous*).”

- (153)
- |            |   |
|------------|---|
| Stage I:   | They are poisonous. That I think.   |
| Stage II:  | They are poisonous, {that I think, I think that /it, as/ so I think}<br>= ‘which I think’ |
| Stage III: | They are poisonous, I think.  |
| OR         |   |
|            | They are poisonous, as I think. = ‘as far as I think, probably’                           |
| Stage IV:  | I think, they are poisonous.<br>They are, I think, poisonous.                             |

The main point in this proposal is that parenthetical instances of *I think* originate in autonomous *I think*-clauses with a pronoun or demonstrative, and which refer to another autonomous clause.

Thompson & Mulac (1991) propose that the epistemic phrase *I think* developed out of a construction in which *I* and *think* are main subject and verb, with *that* introducing a complement clause (ibid:313), as illustrated by Stage I in (154), exemplified with Brinton’s example sentence in (153). According to Thompson & Mulac, instances of *I think* with ‘*that*-deletion’ have been reanalyzed as epistemic phrases (Stage II), or epistemic parentheticals (Stage III). They argue (ibid:326), “[a]s epistemic phrases, then, these combinations are free to float to various positions in the clause [...]” In their view, *that*-deletion is evidence of grammaticalization of epistemic phrases.

- (154)
- |            |   |
|------------|---|
| Stage I:   | <i>I think that they are poisonous.</i> |
| Stage II:  | <i>I think Ø they are poisonous.</i>    |
| Stage III: | <i>They are poisonous, I think.</i>     |

The main point in their proposal is that parenthetical instances of *I think* developed out of complementing structures with *I think*. This synchronic analysis has been criticized for various reasons, but mainly on historical grounds (Aijmer 1997; Brinton 1996; 2008, Fischer 2007; Kearns 2007). Note that sentence-initial *I think*, as in Stage II, is ambiguous between a complementing and parenthetical pattern, (see Brinton 1996; Dehé & Wichmann 2010).

Boye & Harder (2007) relate their analysis of complement taking clauses to both Thompson & Mulac's and Brinton's hypothesis on the development of the mental state predicate *I think*. Boye & Harder distinguish between a structure and usage level, which implies a distinction between 'lexical' and 'grammatical' CTPs (=structural status) and 'primary' and 'secondary' CTPs (=usage status).<sup>85</sup> Lexical CTPs predicate their content of arguments and assign argument roles to them, i.e. they impose their profile on the whole clause. A grammatical CTP is "non-argument-assigning, non-predicating and non-profile-imposing" and functionally similar to epistemic adverbials" (ibid:581). The primary CTP, together with the rest of its CTP clause, expresses the main point of an utterance. The secondary CTP has only a concomitant function in relation to the rest of the utterance (which expresses its main point), (ibid:584).

Boye & Harder (2007:592) reformulate Thompson & Mulac's scenario, adapted to their terminology, as shown in (155).

- (155) [I think] [I love her]  
 A: [main lexical CTP clause, primary status] [complement clause]  
 ↓ usage reanalysis  
 B: [main lexical CTP clause, secondary status] [complement clause]  
 ↓ structure reanalysis, grammaticalization of CTP  
 C: [grammatical CTP clause, secondary status] [main clause]

Both Brinton's scenario in (153), and Thompson & Mulac's idea in (154), are compatible with Boye & Harder's hypothesis in (155), because the main point in all proposals is the development of non-qualificational into qualificational *I think*. This development "must involve an intermediate stage (Brinton's stage II and our stage B) of lexical but secondary CTPs."

Fischer (2007) also compared the ideas of Thompson & Mulac (1991) and Brinton (1996), and comments on the origin of parenthetical phrases like *I think* in English, as quoted in (156), (ibid:112).

- (156) I do not believe they [i.e. parenthetical phrases like *I think* etc.] originally were part of complex clauses as both T[hompson] & M[ulac], and B[rinton] have argued in different ways. They probably occurred both in independent clauses and with complement clauses from the very beginning, the former being most frequent in spoken, the latter in written

---

<sup>85</sup> See also their definition of grammaticalization and grammatical status in (26) and (27) in Chapter 2.

discourse. [...] It seems to me more likely, however, that Gorrell (1895) is right when he accepts the use of parentheticals already for Old English.

To me, this seems to be the most plausible scenario, as there is a wide variety of (overlapping) syntactic structures for *I think* that co-exist these days. Moreover, as noted by Brinton (1996:252), “the link between syntactic and semantic change is hard to establish.” Verbs, in general, figure in various syntactic configurations, so there is no reason to assume that the development of *I think* was restricted to either a complement taking structure or various types of relative clauses.<sup>86</sup> Hence, different clause types, both dependent and independent ones, may (simultaneously) have given rise to the parenthetical variant of *I think*.

The semantic changes in the development of *I think* have only been briefly discussed. Brinton (1996:243) proposes a cline for the semantic development of first person epistemic parentheticals, as in (157).

(157) act of cognition → mode of knowing (evidential) → (un)certainty (epistemic)

At first, mental state predicates denote an act of cognition, which turns into a mode of knowing, which consecutively comes to express degrees of (un)certainty of the proposition.

According to Nuyts (2001:114), “there is a clear link from the concrete mental process to the interpersonal use, which suggests a historical connection between them” (Nuyts 2001:114). The non-qualificational variant of *think* is clearly the literal one and may be considered the original meaning. The epistemic meaning of *think* may have been able to develop out of the non-qualificational meaning because it essentially leaves open the reality-status of the state of affairs. This factual openness appears to have been exploited to indicate uncertainty on behalf of the speaker. This also explains why the epistemic meaning of *I think* is vague and non-specific (ibid:115).

The next section deals with the semantic and syntactic properties of Mainland Scandinavian TROR, its etymology and development. These will be examined in relation to the general observations for *I think* in English as discussed in this section.

### 5.2.2 Mainland Scandinavian TRO and TROR

There is a whole class of mental state predicates in the Mainland Scandinavian languages. Basically, there are three different cognitive verbs in order to convey different aspects of cogitation. These are *tro*, *synes* and *mene* in Danish and Norwegian, and *tro*, *tycka* and *mena* in Swedish, all meaning ‘think/believe’, ‘find’ and ‘mean’ respectively. In this chapter, the

---

<sup>86</sup> Note that these different ideas about the origin of ‘(epistemic) parenthetical phrases like *I think*’ may also be relevant to the development of the epistemic adverb ‘maybe’ which also has a phrasal origin. The developmental scenario of Wessén (1967) resembles Thompson & Mulac (1991) developmental path for *I think*. On the other hand, the variable phrase (*det*) *kan ske (att)*, could also be inserted in various adverbial positions in a clause, which makes a parallel development for complementing and parenthetical patterns plausible for MAYBE as well.

cognitive verb *tro* is compared for Danish, Norwegian and Swedish as it corresponds to English *I think*, and *tro* is a form that is available in all three languages.

The meaning of Mainland Scandinavian TRO (< Old Norse *trúa*, Old Danish *tro*, Old Swedish *troa*, *tro*) is likely to be derived from a Proto-Germanic etymon ‘be firm’ (SAOB, ODS). Nowadays, TRO expresses both relative certainty ‘believe, be almost certain of’ and uncertainty ‘think, assume’.<sup>87</sup> It expresses various qualificational and non-qualificational meanings, and may also occur in idiomatic or fixed expressions. Non-qualificational TRO denotes the concrete mental activity of believing X. Qualificational TRO is a mixed epistemic-evidential form (Nuyts 2001) meaning either ‘believe’ or ‘think’, that is, it express both the source of the information, and relative certainty and uncertainty.

The verb TRO has a full inflectional paradigm in the Mainland Scandinavian languages, see Table 37 below, which also includes present participle imperative and passive forms.

language	infinitive	present	past	perfect
Danish	<i>at tro</i>	<i>tror</i>	<i>troede</i>	<i>troet</i>
Norwegian	<i>å tro</i>	<i>tror</i>	<i>trodde</i>	<i>trodd</i>
Swedish	<i>att tro</i>	<i>tror</i>	<i>trodde</i>	<i>trott</i>

**Table 37.** Inflectional paradigms for Mainland Scandinavian TRO.

The subsections 5.2.2.1, 5.2.2.2, 5.2.2.3 summarize the meanings of *tro* in each of the Mainland Scandinavian languages.

### 5.2.2.1 Danish *tro*

Danish *tro* distinguishes several nuances with respect to belief, thought and doubt. The meanings and corresponding examples, as listed in DDO, are: deem something as probable (assume, believe), as in (158a), have a subjective opinion/idea/assumption about how something is or should be (think, find), as in (158b), feel convinced of something (often in case of wrong assumptions), as in (158c), express uncertainty or carefulness, as in (158d), be personally convinced of something, as in (158e), have faith/trust that something is good or true, as in (158f), have faith/trust in someone’s capacities, feasibility of a project etc., as in (158g), have unconditional religious beliefs, as in (158h), and finally, have a firm belief in the existence of something, as in (158i).

- (158) a. ***Jeg tror, at hun er i København, men jeg kan da få det oplyst.***  
‘I think she is in Copenhagen, but I can get it clarified.’

<sup>87</sup> It is not uncommon that notions expressing certainty come to express uncertainty as well. An example is Dutch *zeker* ‘certain’ or *zeker weten* ‘to know for sure.’ In *Het is zeker dat hij komt* ‘He’ll be there for sure’ *zeker* expresses absolute certainty, but in *Hij komt zeker weer te laat/ Hij is weer te laat, zeker?* ‘He’ll be late again, I guess’ it expresses a probability. Factors such as word order and intonation determine the reading of *zeker*.

- b. *Helt ærligt, **jeg tror**, du skulle finde dig noget billigere at bo i.*<sup>88</sup>  
 ‘To be honest, I think, you should get yourself a cheaper place to live.’
- c. *Han skiftede i Fredericia og **troede sig** i toget på vej til Kolding, da det pludselig gik op for ham, at toget kørte til København.*  
 ‘He changed in Fredericia and thought himself to be in the train towards Kolding, when he suddenly realized, that the train was going to Copenhagen.’
- d. ***Jeg tror godt**, jeg vil have et glas hvidvin.*  
 ‘Yes I think I’d like to have a glass of white wine.’
- e. *Jeg vil se det, før **jeg tror** det.*  
 ‘I want to see it, before I believe it.’
- f. ***Jeg tror på** det, min mor fortalte.*  
 ‘I believe in the things my mother told.’
- g. *Britt Nørlem **tror på sig selv** og sin butik.*  
 ‘Britt Nørlem believes in herself and her shop.’
- h. *Jesus siger: “Den, som **tror** og bliver døbt, skal blive frelst; men den, som er vantro, skal blive fordømt.”*  
 ‘Jesus says: ‘S/he who believes and gets baptized, will be saved; but s/he who is unbelieving, will be condemned.’
- i. *Ingen af os to **tror vel på** genfærd eller spøgelser.*  
 ‘None of us does believe in spirits or ghosts, do we?’

Some idiomatic expression with *tro* are *det kan du tro* ‘that I ensure, obviously’, *det tror pokker/fanden* which is strong language for ‘obviously’ and *tro om (igen)* ‘reconsider.’ Note that only (158a,b) express qualificational meanings.

### 5.2.2.2 Norwegian *tro*

Norwegian *tro* ‘believe, think’ expresses various dimension of certainty and uncertainty, as listed in BO, and cited in (159) below. It may mean ‘be almost certain of something, assume, believe, think’, as in (159a). It may also express that one has faith or trust in something, as in (159b), which also has a reflexive variant (159c), or that one is convinced of something, as in (159d).

<sup>88</sup> This clause is grammatically correct, but the mental state predicate *jeg synes* ‘I find’ would be more appropriate in this example because it concerns a personal opinion and not a statement one is not sure about.

The infinitival form *tro* may also occur as a question particle to mark doubt, in these cases *tro* is a short form of the expression *mon tro* or *skal tro*, as in (159e), see also Section 4.4 on MON.

- (159) a. ***Jeg tror det blir regn***  
 ‘I think it will rain.’
- b. *En skal ikke tro alt en hører.*  
 ‘One should not believe everything one hears.’
- c. *Han tror seg ikke til en så vanskelig oppgave.*  
 ‘He does not commit himself to such a difficult task.’
- d. *Han tror på Gud.*  
 ‘He believes in God.’
- e. *Blir det regn, tro?*  
 ‘Will there be rain, do you think?’

Only the examples in (159a,e) convey qualificational meaning. Example (159a) is ambiguous between a complementing pattern with *that*-deletion and a parenthetical pattern with sentence-initial *I think*.

### 5.2.2.3 Swedish *tro*

Swedish *tro* expresses a variety of qualificational and non-qualificational meanings related to thoughts and beliefs, as listed in NSSO, and quoted in (160). These include, deem something as probable (assume, believe), as in (160a), have faith/trust in someone’s capacities, as in (160b), mitigating expression about (undetermined) intentions, as in (160c), believe/hold something for being true, as in (160d), believe someone, take someone at his/her word, as in (160e), have faith/trust in someone’s capacities, feasibility of a project etc., as in (160f), have unconditional religious beliefs, as in (160g), adverbial *tro* may be used as a question marker, as in (160h), cf. Chapter 4 on MON.

- (160) a. ***Jag tror att det blir regn.***  
 ‘I think/believe there will be rain.’
- b. *Han trodde sig kunna gjennomføre oppgiften.*  
 ‘He thought he could complete the task.’
- c. ***Jag tror (att) jag väntar en stund.***  
 ‘I think (that) I will wait a little bit’

- d. *Man skall inte **tro** (på) allt som man hör ryktesvis.*  
'One should not believe (in) everything,  
one hears through the grapevine.'
- e. *Jag **tror** (på) honom.*  
'I believe (in) him.'
- f. *Chefen **trrodde på** honom och gav honom allt mer ansvarsfyllda uppgifter.*  
'The boss believed in him and gave him more and more responsible tasks.'
- g. *"Tvivla inte utan **tro**!" sade Jesus.*  
'Do not doubt but believe! said Jesus'
- h. *Kommer han ikväll, **tro**?*  
'Is he coming this evening, I wonder.'

The cognitive verb *tro* also occurs in a number of idiomatic expressions. It may be used as a purely reinforcing expressions *du kan inte tro vad roligt vi hade!* 'you won't believe what a good time we had' or *det tror jag det!* 'of course that's the case', to express mild doubt *månn tro (det)*, 'really?' or doubtful astonishment with *tro det!*

#### 5.2.2.4 Mainland Scandinavian TROR

Mainland Scandinavian TROR has both non-qualificational (=believe (in) *X*) and qualificational (=assume + speech-act) meanings in Danish, Norwegian and Swedish, as described in 5.2.2.1-3 above. Besides this semantic variation, there is a wide variety of syntactic structures available for TROR. It may be a lexical verb in a main clause, it may be a complement taking predicate, or it may be a discourse marker. All of these different functions correspond to different syntactic manifestations.

In the literature it has often been observed that mental state predicates like *I think* have an adverbial distribution. In this section, I will examine some of the examples in 5.2.2.1-3, in order to find out to what extent Mainland Scandinavian TROR has acquired adverbial status.<sup>89</sup> More specifically, qualificational examples with a first-person subject will be placed in Platzack's joint sentence scheme. The attested structures, as well as the structural status for TROR, will then be compared to the epistemic adverb MAYBE<sup>90</sup> and the speech-act adverbial *ärligt talat* 'honestly' in Swedish (Aijmer 1997).

The structural status of TROR can be analyzed in three different ways, namely as a lexical verb in a main clause (MC), a complement taking predicate (CTP), or as a discourse markers

<sup>89</sup> Note that *tro* also has given rise to univerbated adverbs such as *kanstro*, *månnstro*, *måtro* in Swedish, but there are also derived adverbs in all Mainland Scandinavian languages like *trolig*, *troligen*, and *troligvis*. Interestingly, *tro* often occurs in expressions that contain a modal auxiliary, e.g. *kan jeg tro* or *må du tro*.

<sup>90</sup> The development of MAYBE has a lot in common with I THINK (cf. Boye & Harder 2007).



(DM). In order to accommodate discourse markers in the scheme, it has to be extended with a ‘prefoundation’-field (PF). The complex structural status of TROR is complicated by the fact that there are no overt formal clues that signal MC, CTP or DM status. In addition, instances of TROR may be ambiguous between these different statuses.

The Swedish example in Table 38 contains an overt complementizer, which makes it ambiguous between a MC and CTP. Because *jag tror* conveys an epistemic judgment, i.e. a qualificational meaning, its status is best captured by a complement taking structure.

	F	T	s	a	V	S	A
MC	<i>Jag</i>	<i>tror</i>				<i>att det blir regn.</i>	
	I	believe				<i>that it will rain.</i>	
CTP	<i>Jag</i>	<i>tror</i>					
	I	think					
		<i>att</i>	<i>det</i>		<i>blir</i>	<i>regn.</i>	
		that	it		will	rain.	

**Table 38.** The structural representation of MC and CTP status for TROR.

The Norwegian example in Table 39 is similar to the Swedish one in Table 38, the only difference is that there is no (overt) complementizer. This example is ambiguous between CTP and DM status. In order to be placed in the prefoundation, there must have been a reanalysis from VP to ADV (as represented by the square brackets).

	PF	F	T	s	a	V	S	A
CTP		<i>Jeg</i>	<i>tror</i>					
		I	think					
			(at)	<i>det</i>		<i>blir</i>	<i>regn.</i>	
			(that)	it		will	rain.	
DM	[ <i>Jeg tror</i> ]	<i>det</i>	<i>blir</i>	<i>regn.</i>				
	[I think]	it	will	rain.				

**Table 39.** The structural representation of CTP and DM status for TROR.

In the literature, English *I think* has been characterized as an epistemic parenthetical. The structural representations of Mainland Scandinavian TROR above show that it may be ambiguous between MC, CTP and DM status in the absence of formal clues that uniquely characterize these different statuses. In order to get a clearer picture of its structural status, TROR will be compared to the epistemic adverb MAYBE and the speech-act adverbial HONESTLY/TO BE HONEST.

The development of MAYBE and TROR have a lot in common. Both have a complementing and parenthetical variant, and they primarily express qualificational meanings. In this sense, MAYBE is the impersonal variant (It may be that X > Maybe X), and I THINK the

personal variant (I think that X > I think, X/ X, I think). In order to find out whether TROR can occur in canonical positions for sentence adverbs, it will be compared with the epistemic sentence adverb MAYBE in Table 40 below.

	F	T	s	a	V	S	A
a.	<i>Han</i>	<i>kommer</i>		<i>kanske inte</i>			<i>idag.</i>
	He	comes		maybe not			today.
*	<i>Han</i>	<i>kommer</i>		<i>[jag tror] inte</i>			<i>idag.</i>
	He	comes		[I think] not			today.
b.	<i>Kanske</i>	<i>kommer</i>	<i>han</i>	<i>inte</i>			<i>idag.</i>
	Maybe	comes	he	not			today
*	<i>[Jag tror]</i>	<i>kommer</i>	<i>han</i>	<i>inte</i>			<i>idag.</i>
	[I think]	comes	he	not			today.
c.	<i>Han</i>	<i>kanske</i>		<i>inte</i>	<i>kommer</i>		<i>idag.</i>
	He	maybe		not	comes		today.
*	<i>Han</i>	<i>[jag tror]</i>		<i>inte</i>	<i>kommer</i>		<i>idag.</i>
	He	[I think]		not	comes		today.
d.	<i>Kanske</i>	<i>(att)</i>	<i>han</i>	<i>inte</i>	<i>kommer</i>		<i>idag</i>
	Maybe	(that)	he	not	comes		today.
*	<i>[Jag tror]</i>	<i>(att)</i>	<i>han</i>	<i>inte</i>	<i>kommer</i>		<i>idag.</i>
	[I think]	(that)	he	not	comes		today.

**Table 40.** Positions for sentence adverbs: MAYBE vs TROR.

Mainland Scandinavian TROR has not acquired the status of an epistemic adverb because it is well-integrated in the syntactic structure and imposes subordinate structure on the clause, compare Table 38-Table 40. Confusion about its structural status probably arises from the fact that TROR may have parenthetical status, and as such functions as an insertion, afterthought or personal comment in the prefoundation field.

Aijmer (1997) notes that word order rules may distinguish speech-act adverbials from other adverbials. Since the Mainland Scandinavian languages are V2 languages, initial sentence adverbials should cause subject-verb inversion. This rule is only optional for initial speech-act adverbials (e.g. *ärligt talat* ‘to be honest, honestly’), but obligatory for other sentence adverbials (e.g. *förmodligen* ‘presumably’), as the examples from Aijmer (1997:4) in Table 41 show.

As regards TROR, which cannot convey personal opinions (these are expressed by *synes* and *tycka*), it cannot be substituted with *ärligt talat* or *förmodligen*. Example c is an instance of a mitigating expression about (undetermined) intentions, which may be located in the prefoundation field, but because *that*-complementation is possible, it is ambiguous between CTP and DM status.

	PF	F	T	s	a	V	S	A
a.	<i>Ärligt talat,</i>	<i>han</i>	<i>är</i>		<i>inte vidare</i>		<i>trevlig.</i>	
	Frankly spoken,	he	is		not very		nice.	
		<i>Ärligt talat</i>	<i>är</i>	<i>han</i>	<i>inte vidare</i>		<i>trevlig.</i>	
		Frankly spoken	is	he	not very		nice.	
b.	<i>*Förmodligen</i>	<i>han</i>	<i>är</i>		<i>inte vidare</i>		<i>trevlig.</i>	
	*Probably	he	is		no very		nice.	
		<i>Förmodligen</i>	<i>är</i>	<i>han</i>	<i>inte vidare</i>		<i>trevlig.</i>	
		Probably	is	he	not very		nice.	
c.	<i>[Jag tror]</i>	<i>jag</i>	<i>väntar</i>				<i>en stund.</i>	
	I think	I	wait				a little	
		<i>Jag</i>	<i>tror</i>					
			<i>att</i>	<i>jag</i>		<i>vän- tar</i>	<i>en stund.</i>	

**Table 41.** Positions for speech-act adverbials: HONESTLY vs TROR.

Mainland Scandinavian TROR has a mitigating function which may be analyzed as an independent syntactic unit that does not affect the structure of the clause to which it attaches. As will become clear in the next sections, TROR may also occur as an insertion or afterthought. The meaning and structure of TROR are closely intertwined. Only in case TROR conveys intention/mitigation, or an additional comment to a clause, it may be situated in the prefoundation-field, or be an insertion or afterthought. In case TROR means ‘believe (in) X’ it is a lexical verb in a main clause. As a mixed epistemic/evidential form it has CTP status as it imposes a subordinate profile on its complement.

### 5.3 Sources and method

The data for the case study of TROR have been selected from the online corpora *KorpusDK*, *Norsk Aviskorpus* and *Språkbankens konkordanser*, cf. Section 1.4. As mentioned in Section 5.2, discourse markers are predominantly a feature of spoken language rather than written discourse. However, for reasons of consistency, the same corpora of written data have been used to select data for all the case studies in the present study. An investigation of TROR in written data may not coincide with identified tendencies in spoken data, but it provides nonetheless information about its use in written contexts. In written corpora, depending on genre, the majority of occurrences of TROR will be descriptive. Few performative uses of TROR are to be expected, as writing is about reporting in an objective way in case of newspaper texts or scientific articles. In spoken language the majority of instances of TROR is performative (cf. Nuyts 2001), but newspaper texts do contain a lot of quotations of people’s opinions on various matters.

Random samples of 1000 occurrences of the form *tror* (present tense of *tro*, which is the

same for all persons), have been taken for Danish, Norwegian and Swedish. To get an overall picture of the development and distribution of TROR, *tror* is left unspecified for person, clausal position (initial, medial or final) and *that*-deletion in order to see whether the phrase is indeed most common as a flexible unit with a first person pronoun. Only the present tense is taken into account as this is the prototypical form of qualificational mental state predicates. The data are analyzed according to a set of criteria, as described in Section 5.2, in order to examine the structural status of TROR.

The first variable is *meaning*. The ratio for non-qualificational meaning (MC) and qualificational meanings (CTP + DM) will be examined as qualificational meanings should be dominant in case *I think* is developing into a discourse marker. The second variable is *person*. In case *I think* is developing into a discourse marker one would expect predominance of first person subjects. The third variable is *that-deletion* (cf. Thompson & Mulac 1991). Synchronically, this is a relevant factor because TROR as a discourse marker cannot be followed by a *that*-clause, i.e. as an independent syntactic unit it does not affect the structure of the proposition. The fourth variable is *position*. As a discourse marker (or sentence adverb), TROR should have a variable distribution with regard to clausal positions, i.e. clause-initial, -medial (insertion) or -final position (afterthought).

The ratings for these variables are presented, discussed and compared for Danish, Norwegian and Swedish *tror* in Section 5.4.

## 5.4 Results

This section presents the results of a comparative corpus investigation of TROR in the Mainland Scandinavian languages. Instances of TROR are examined with respect to the variables listed in (161) below

- (161) a. *meaning* (non-qualificational or qualificational)  
 b. *person* (first, second or third person subject)  
 c. *that-deletion* (complementizer or no complementizer)  
 d. *position* (clause-initial, -medial or -final)

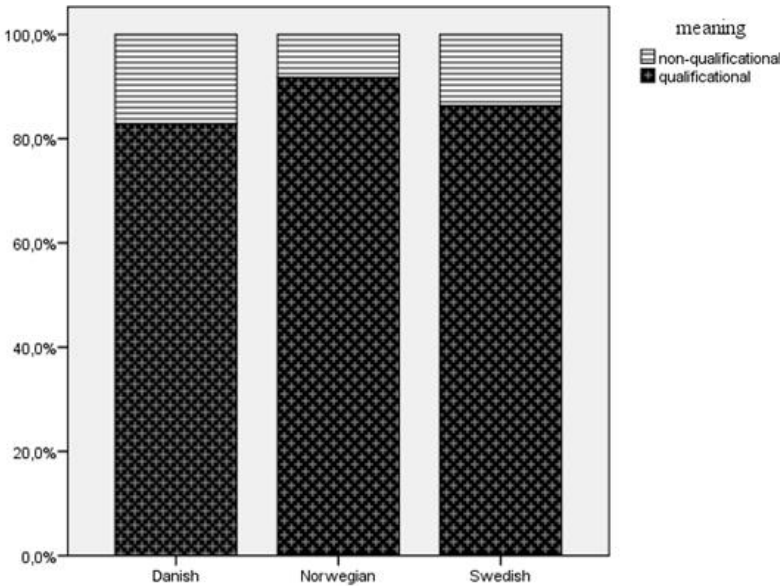
These distributional properties are discussed and illustrated by means of corpus examples in the Sections 5.4.1-5.4.7 below.

### 5.4.1 Qualificational and non-qualificational meanings

The cognitive verb TROR conveys a variety of different meanings associated with ‘belief’, ‘opinion’ and ‘intention’ cf. also section 5.2.2. In case TROR is moving towards (primarily) parenthetical status, qualificational meanings should be dominant. The ratio for qualificational and non-qualificational meanings distinguishes between lexical and epistemic/evidential and communicative uses of TROR.

As shown in Figure 9 below, qualificational meanings are by far most frequent ones for

TROR. They constitute 82.9 %, 91.7 % and 86.3 % of the Danish, Norwegian and Swedish samples respectively. The corresponding percentages for non-qualificational meanings are 17.1%, 8.3 % and 13.7% respectively. There is a significant difference with respect to the distribution of qualificational and non-qualificational meanings in the Mainland Scandinavian samples ( $\chi^2=34.7$ ,  $df=2$ ,  $p<0.01$ , Cramer's  $V=0.108$ ). See Appendix 5 for the counts per category in each language.



**Figure 9.** Bar chart of the distribution for (non-)qualificational meanings of TROR.

As described in more detail in Section 5.2, non-qualificational meanings of TROR include various nuances of belief. They are transitive, lexical verbs in declarative main clauses, as shown by the corpus examples in (162) below.

- (162) a. *Hvis du ikke **tror mig**, så læs teksterne til singlen " Gett Off ",* Danish  
*og check om dine ører ikke bli'r helt røde?*  
 'If you don't believe me, then read the lyrics of the single  
 'Gett Off' and check if your ears won't get red.'
- b. *De **tror på** oss, og det setter vi pris på,* Norwegian  
*sier Fjeldheim til Aftenbladet.*  
 'They believe in us, and we appreciate that,  
 says Fjeldheim to Aftenbladet.'

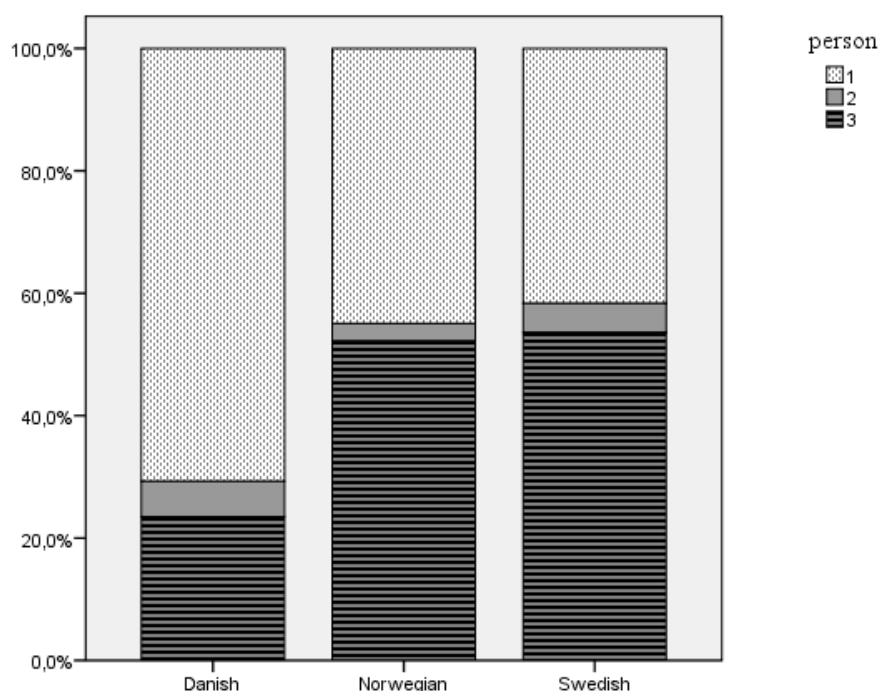
- c. *Men till slut **tror** jag honom.*  
 ‘But in the end I believe him.’

Swedish

These non-qualificational meanings will not be discussed any further in this study as they do not develop into discourse markers. In the next section, the qualificational instances of TROR will be discussed with respect to the variable person.

### 5.4.2 Person

In case Mainland Scandinavian TROR develops into a discourse marker one would expect predominance of a first person subject. Figure 10 below shows that Danish *tror* predominantly occurs with a first person pronoun, whereas Norwegian and Swedish *tror* predominantly occur with third person subjects. This is a significant difference ( $\chi^2=204.8$ ,  $df=5$   $p<0.01$ , Cramer’s  $V=0.197$ ). See Appendix 6 for an overview of the counts per category in each language.



**Figure 10.** Bart chart of the distribution for first, second and third person subjects with TROR.

Performative uses of TROR, i.e. with a first person subject, constitute 70.7 % of the Danish sample, and 45.0 % and 41.7 % of the Norwegian and Swedish samples respectively. Corpus examples of performative uses of Mainland Scandinavian TROR are given in (163) below.

- (163) a. ***Jeg tror** kritikken afspejler en frygt for egen fremtid.* Danish  
 ‘I think the criticism reflects a fear of one’s own future.’
- b. *Å være seg selv, det **tror jeg** er viktig.* Norwegian  
 ‘To be yourself, I think that is important.’
- c. *Vid Gullmarsplan ska man byta till linje 18, **tror jag**.* Swedish  
 ‘At Gullmarsplan one has to change to number 18, I think.’

Second person pronouns occur marginally with TROR in written contexts, only 5.8 %, 2.9 % and 4.7 % of the Danish, Norwegian and Swedish samples respectively. This combination is generally used to address someone, and is most often a question format. Its infrequency can be explained by the fact that these interactive uses are more typical of conversations than newspaper texts which are clearly not an interactive medium. For these reasons, TROR in combination with a second person pronoun will not be further commented on in this study. Corpus examples of TROR with a second person subject are given in (164) below.

- (164) a. *Det er ikke som **du tror**.* Danish  
 ‘It is not what you think’
- b. *Hva **tror du**?* Norwegian  
 ‘What do you think?’
- c. ***Tror du att** du ska komma till ett annat slags liv,  
 är det vad **du tror**?* Swedish  
 ‘Do you think/believe that you will get another kind of life,  
 is that what you think?’

Descriptive and ascriptive uses of TROR are most frequent for Norwegian and Swedish TROR, 52.2 % and 53.6 % respectively, but occur much less, 23.5 %, in the Danish sample. Corpus examples of descriptive/ascriptive TROR are given in (165).

- (165) a. ***Politiet tror**, at ofret selv har lukket gerningsmanden ind.* Danish  
 ‘The police think that the victim has locked in the perpetrator.’

- b. *Selv **tror han** at det hele var en ulykke.* Norwegian  
 ‘Personally he thinks that it was all an accident.’
- c. *Ett vulkanutbrott under vannet, **tror** några.* Swedish  
 ‘A volcanic eruption under water, some think.’

In the remainder of this chapter, performative and descriptive/ascriptive uses TROR will be discussed with regard to *that*-deletion and position. The main focus will be on first person pronoun combinations, the third person pronoun combinations with TROR will not be discussed in great detail, but serve as a means of comparison. For the performative uses, which have the potential to develop into a discourse marker, it is expected that they are moving towards an adverbial distribution, whereas the descriptive/ascriptive uses are expected to retain verbal distributions.

### 5.5.3 *That*-deletion

The relation between main and subordinate clause is looser in case the complementizer is omitted. This may be seen as a sign of adverbialization. If the phrase TROR is acquiring adverbial status, one should expect the complementizer to disappear because (sentence) adverbs are generally not followed by a complementizer, e.g. \*possibly *that*.<sup>91</sup>

The distribution of clauses with and without *that* differs considerably in the Mainland Scandinavian languages. Swedish *tror* is much more often followed by *that* (61.9%) than in Norwegian (9.4%) or Danish (32.4%). Norwegian *tror* generally occurs without *that* (90.6%), in Danish and Swedish *that*-less clauses constitute 67.6 % and 38.1 % of the samples respectively.

Recall that ‘that’ is an optional complementizer (cf. Chapter 4). I have no explanation as to why Swedish *tror* primarily co-occurs with *that*, whereas instances of Norwegian *tror* are characterized by *that*-deletion. It might be a language specific preference, but more research is needed on the frequency and the optionality of *that* in the Mainland Scandinavian languages. This goes beyond the scope of the present thesis, as the presence or absence of *that* is used as a factor in adverbialization. Examples with a *that*-complement are instances of CTP structures, examples with *that*-deletion are ambiguous between CTP and DM. However, note that ‘that’ is possible in the *that*-less clauses and that DM status is only possible with communicative meanings (i.e. not with epistemic/evidential ones).

Corpus examples of TROR with a *that*-complement are given in (166) below.

- (166) a. ***Jeg tror at** tiden er kommet til at pendulet skal svinge tilbage* Danish  
*til de mere menneskelige værdier, siger Kirsten Hastrup.*  
 ‘I think that the time has come that the pendulum will swing  
 back to the more human merits, Kirsten Hastrup says.’

<sup>91</sup> But see Chapter 4 on MAYBE which may optionally be followed by the complementizer *that*, and various types of insubordination phenomena.

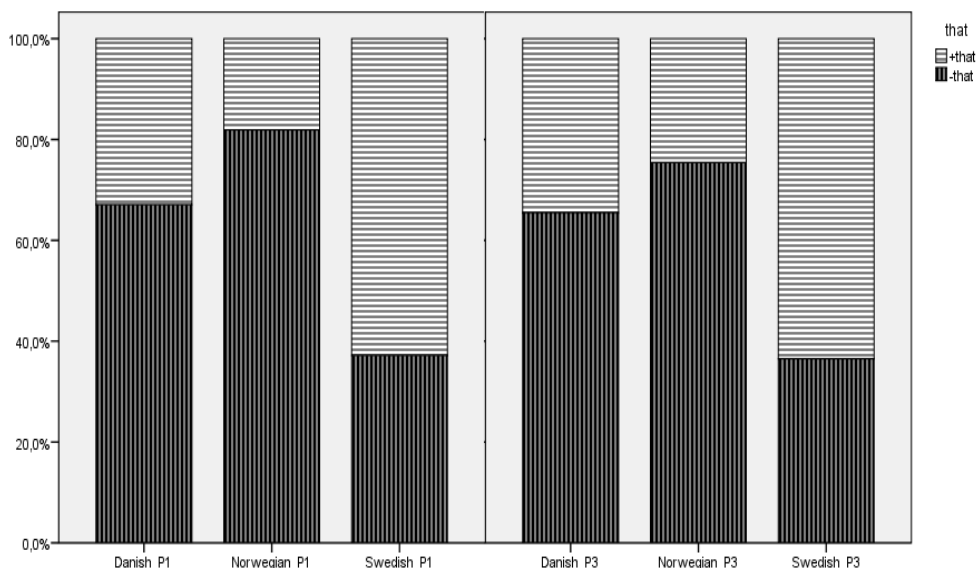


- b. **Jeg tror at** når en dør lukkes, er det en ny som åpner seg. Norwegian  
 ‘I think that when a door closes, there will be a new one that opens.’
- c. **Jag tror att** alla vet att vi är "nya" här i kväll. Swedish  
 ‘I think that everyone knows that we are ‘new’ here this evening.’

Corpus examples of TROR and *that*-deletion are given in (167).

- (167) a. **Jeg tror** det var et kviksand, han kom ind i. Danish  
 ‘I think it was quicksand he sank into.’
- b. **Jeg tror** litt av det som har skjedd er tilfeldig. Norwegian  
 ‘I think little of what has happened was coincidental.’
- c. **Jag tror** vi har en bra chans. Swedish  
 ‘I think we stand a good chance.’

The ratio for *that*-complementation and *that*-deletion for first person subjects (P1) and third person subjects (P3) is shown in Figure 11 below. See Appendix 7 for the counts per category in each language.



**Figure 11.** Bar chart of the distribution for *that*-complementation and *that*-deletion for TROR.

Since this bar chart contains three variables (*language*, *person* and *that-deletion*), and is thus more complex, a Chi-square test cannot be used to analyze these data. Instead, *Loglinear Analysis*, which is an extension of Chi-square, may be applied. Loglinear Analysis (e.g. Field 2005:695-720) provides a robust analysis of complicated contingency tables and is used to describe associations and interaction patterns among a set of categorical variables.

The procedure is that hierarchical Loglinear models are fitted to multi-dimensional cross tabulations with the aim to find the best model for describing relationships between categorical variables. The analysis starts with a so-called 'saturated model' that contains all interactions and main effects. Backward elimination is used to find out whether deletion of a predictor (variable or combination of variables) has a significant effect on the explanatory power of the model.

The technique is hierarchical in the sense that it starts with removing the highest-order interaction (*language\*person\*that*), next the lower-order, two-way, interactions (*language\*person*, *language\*that-deletion* and *person\*that-deletion*) will be deleted from the model, and finally the main effects of the individual variables (*language*, *person* and *that-deletion*) will be eliminated. In case deletion of the highest-order interaction is significant the analysis stops. There is no use in analyzing the lower-order interactions and/or main effects any further as these effects are all confounded with the highest, three-way, -order interaction.

A disadvantage of *Loglinear Analysis* is that one cannot identify dependent and independent variables. In this case study, for example, it would be desirable to specify predictors that can explain the distribution of a given variable. Moreover, in order to find out which interactions and/or main effects contribute to the significance of a three-way interaction, one still has to 'break down' the table and carry out separate Chi-square tests on these variables.

So in order to analyze all of the data as illustrated by the bar chart in Figure 11, a Loglinear Analysis and separate Chi-square tests had to be carried out. The Loglinear Analysis shows a significant three-way interaction effect for *language\*person\*±that* ( $\chi^2=20.37$ ,  $df=2$ ,  $p<0.01$ ). To break down this effect, separate Chi-square tests on the distribution of first and third person subjects in combination with *that-deletion* have been carried out.

There is a significant difference for TROR and the presence or absence of the complementizer *that* in Danish, Norwegian and Swedish ( $\chi^2=206.8$ ,  $df=2$ ,  $p<0.01$  and Cramer's V 0.412), i.e. *language\*±that*. What one would expect in case TROR is becoming a discourse marker is that first person subjects predominantly figure in *that*-less clauses as they are acquiring an adverbial distribution, unlike clauses with third person subjects. For Danish and Swedish there is no difference between first and third person subjects and the presence or absence of *that*. But for Norwegian there is a significant difference between first and third person subjects and the presence or absence of *that* ( $\chi^2=28.97$ ,  $df=1$ ,  $p<0.01$ ).

### 5.5.5 Position

The mental state predicate TROR may occur clause-initially, -medially or -finally. The latter two may be instances of parenthetical use, i.e. insertions or afterthoughts. Clause-initial instances may be ambiguous between a complementing and parenthetical pattern. Most instances of

clause-medial TROR are due to topicalization. Matrix clauses, in which a constituent is preposed or topicalized, are characterized by inversion of the subject and verb, but these do generally not qualify as parentheticals. Some corpus examples of clause-medial TROR are given in (168) below.

- (168) a. *Opmærksomhed* **tror** **jeg** *er et af* Danish  
 Awareness think I is one of  
*kodeordene i denne tid.*  
 keywords-the in this time.  
 ‘Awareness is, I think, one of the keywords these days.’
- b. *Konsekvensene* **tror** **jeg** *du* Norwegian  
 Consequences.the think I you  
*hadde villet annerledes.*  
 Had wanted otherwise  
 ‘The consequences I think you would have liked to have had the consequences otherwise.’
- c. *Den brasilianska gruppen från Bahia* **tror** **jag** Swedish  
 The Brazilian group.the from Bahia think I  
*blir något alldeles särskilt, säger han.*  
 will something very special, says he.  
 ‘The Brazilian group from Bahia I think will be something very special, he says.’

Unambiguous parenthetical instances of TROR occur rarely in the samples, but there are a few instances in the corpus data. Note that TROR in clause-initial position may occasionally be omitted, strictly speaking these clause-initial non-inverted strings are not parentheticals as they do not occur clause-medially or –finally. Nonetheless, they behave like discourse markers as they do not convey an epistemic/evidential dimension, but express an intention or function as a mitigator, as in (169).

- (169) a. **Jeg tror, jeg er bange.** Danish  
 ‘I think, I am scared.’
- b. **Jeg tror jeg skal skifte telefonnummer.** Norwegian  
 ‘I think I will change my phone number.’

- c. *Jag tror jag lutar mej tillbaka en stund och vilar ögonen,* Swedish  
*väck mej om det händer nåt...*  
 ‘I think I will lean backwards for a moment and rest my eyes,  
 wake me up if something happens...’

Parenthetical instances of TROR occur rarely in the samples, but there are a few instances in the corpus data. Corpus examples of parenthetical instances of TROR in clause-medial position are given in (170) below.

- (170) a. *På hjørnet er der en vogn, der sælger mad, frankfurtere* Danish  
*eller pølser, **tror jeg**, og sodavand.*  
 ‘At the corner there is a stall that sells food, sausages I think,  
 and soft drinks.’
- b. *Min styrka, **tror jag**, ligger i att jag kan entusiasmera laget.* Swedish  
 ‘My strength, I think, lies in that I can excite the team.’

Note that parentheticals do not necessarily have to be inversions of the subject and finite verb, see (171).

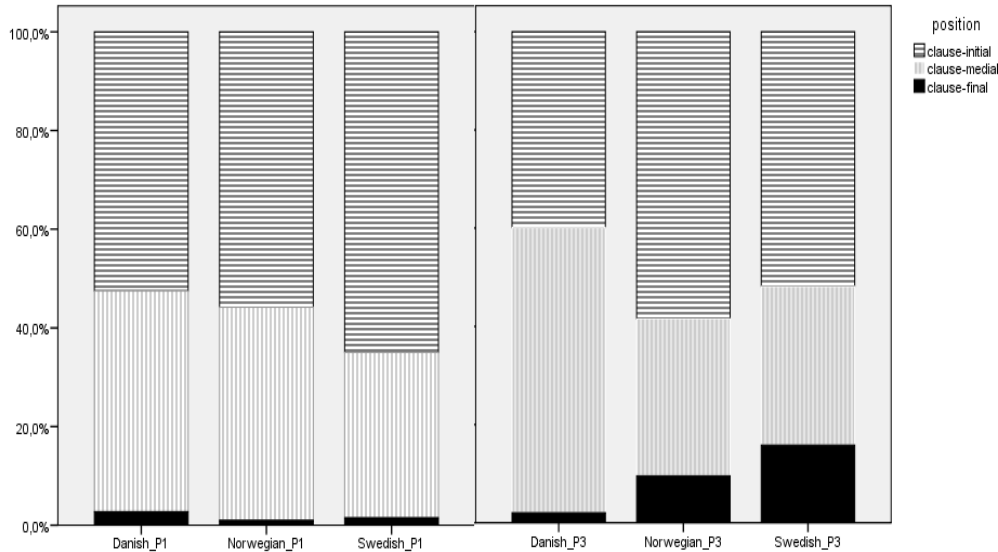
- (171) a. *Han skal ikke se drageagtig ud, **jeg tror**, han skal spille det.* Danish  
 ‘He should not look dragon-like, I think, he should play it.’
- b. *Det har resulterat i att sysselsättningen ökat med, **jag tror**,* Swedish  
*45 000 arbetstillfällen.*  
 ‘It has resulted in that the employment increased with, I think,  
 45,000 jobs.’

Corpus examples of parenthetical instances of TROR in clause-final position, are given in (172) below.

- (172) a. *Desværre gjorde det ikke indtryk på Agathe, snarere tværtimod **tror jeg**.* Danish  
 ‘Unfortunately it did not impress Agathe, rather the opposite I guess.’
- b. *En ting er jeg gammel nok til å forstå dynamikken i, **tror jeg**.* Norwegian  
 ‘There is one think I am old enough to understand the dynamics  
 of, I think.’
- c. *Det handlar om att inte uppleva någon inre glädje, **tror jag**.* Swedish  
 ‘It is about not experiencing any inner happiness, I think.’

Instance of clause-initial TROR (both complementing and parenthetical) constitute 52.5% of the Danish sample, 55.8 % of the Norwegian sample and 65% of the Swedish sample. These percentages are 44.8%, 43.2% and 33.5% for clause-medial position and 2.7%, 1% and 1.5 % for clause-final position, respectively. The distribution of the different clausal positions for first person subjects (P1) is shown in Figure 12 below. The distribution for third person subjects (P3) is presented for comparison.

There is a significant three-way interaction effect for the variables *language\*person\*position* ( $\chi^2=35.5$ ,  $df=8$ ,  $p<0.01$ ). The position of TROR differs significantly in Danish, Norwegian and Swedish ( $\chi^2=15.9$ ,  $df=4$ ,  $p<0.01$  and Cramer's  $V=0.082$ ), i.e. *language\*position*.



**Figure 12.** Bar chart of the distribution for clause-initial, -medial and -final TROR.

The differences per language for position and third and first person subjects are all significant. The results of the separate Chi-square tests are for Danish ( $\chi^2=10.02$ ,  $df=2$ ,  $p<0.01$ , Cramer's  $V=0.116$ ), Norwegian ( $\chi^2=28.16$ ,  $df=2$ ,  $p<0.01$ , Cramer's  $V=0.192$ ) and Swedish ( $\chi^2=45.5$ ,  $df=2$ ,  $p<0.01$ , Cramer's  $V=0.245$ ).

### 5.4.7 Overall results

The general picture that arises from the random samples is that there is a lot of syntactic and semantic variation for TROR, which suggests that this mental state predicate is in full development. Despite significant differences between Danish, Norwegian and Swedish with respect to the distributional properties of TROR, it is hard to detect clear tendencies towards

lexicalization, grammaticalization or pragmaticalization (cf. Section 5.5), or even to identify different developmental stages in the Mainland Scandinavian languages. In all three languages the predominant meanings of TROR are qualificational ones. First person subjects are more frequent in the Danish sample than in the Norwegian and Swedish samples. Swedish *tror* is usually followed by a *that*-complement, whereas Norwegian *tror* is characterized by *that*-deletion. The canonical position for Mainland Scandinavian TROR is the clause-initial position and clause-medial positions come second. The samples contain a handful instances of clear-cut parentheticals, the vast majority of the data show structural ambiguity between MC, CTP and DM status.

## **5.5 Mainland Scandinavian TROR: lexicalization, grammaticalization or pragmaticalization?**

As we have seen in the previous sections, Mainland Scandinavian TROR may assume different functions, which is reflected by its structural ambiguity. In what follows, I will discuss and analyze the structural status of TROR in relation to the converging and diverging properties of lexicalization, grammaticalization and pragmaticalization and (inter)subjectification.

First the mechanisms in lexicalization, grammaticalization and pragmaticalization will be discussed in subsection 5.5.1, then the accompanying primitive changes will be examined in subsection 5.5.2, the possible side effects are elaborated on in subsection 5.5.3, and finally, in 5.5.4 I will comment on the linguistic status (i.e. lexical, grammatical or communicative) and degrees of lexicalization, grammaticalization or pragmaticalization, and (inter)subjectification for Mainland Scandinavian TROR.

### **5.5.1 Mechanisms in the development of TROR**

As defined in the present study, lexicalization, grammaticalization and pragmaticalization are essentially conceived of as the result of formal reanalysis and semantic reinterpretation. With regard to TROR, the scheme looks as in Table 42 below. The cells that apply to TROR are shaded.

The complex structural status of TROR, which at first sight cannot be formally distinguished, relates to the different meanings and functions that it may convey. Mainland Scandinavian TROR may express i) belief (in) X, ii) assumption, and iii) intention/mitigation, or a combination of these. As a lexical verb meaning ‘believe (in) X’, sentence-initial TROR is a declarative main clause. When sentence-initial TROR conveys assumption, it is a mixed epistemic/evidential form (i.e. an assumption rooted in the speaker’s beliefs about a certain state of affairs) in a complement taking pattern. In case sentence-initial, -medial or -final TROR expresses intention, or purely communicative meanings, it is a parenthetical, which is a loosely attached comment to the propositional content of the clause.

As described in Section 5.2.1, the origin of phrasal discourse markers like ‘I think’ is unclear. Various dependent and independent clause types may simultaneously have given rise to the parenthetical instances of ‘I think’. In the development of TROR there has been a hierarchical reanalysis from propositional to extra-propositional status. Constituent internal reanalysis relates

to the extent to which TROR is a (non-fused) unit [JEG TROR]/[TROR JEG] or a phrase [JEG[TROR]]. As a discourse marker, it is a syntactically independent unit, attached to the proposition. There are contiguous semantic relations between the cognitive verb that denotes an act of cognition (believing (in) X), which turns into a mode of knowing (belief, supposition), which then came to express degrees of (un)certainty of the proposition (I am not entirely sure), as well as opinion.

<b>i. Mechanisms in language change</b>	<b>Lxn1</b>	<b>Lxn2</b>	<b>Gzn1</b>	<b>Gzn2</b>	<b>Pgzn</b>
<b>reanalysis</b>					
<b>- hierarchical reanalysis</b>					
° propositional > extra-propositional status	-	-	-	-	+
<b>-categorical reanalysis</b>					
° major > minor category	-	-	+	-	±
° minor > minor category	-	-	-	+	±
<b>-constituent internal reanalysis</b>					
° syntagm/complex lexeme > (simple) lexeme	+	-	(+)	(+)	(+)
° bound morpheme > semi-independent word	-	+	-	-	-
<b>reinterpretation</b>					
<b>-metaphor/metonymy</b>					
° referential > referential meaning	+	-	-	-	-
° referential > relational meaning	-	-	+	-	-
° relational > relational meaning	-	-	-	+	-
° referential/relational > referential meaning	-	+	-	-	-
° referential/relational > communicative meaning	-	-	-	-	+

**Table 42.** Mechanisms in the development of TROR.

### 5.5.2 Primitive changes in the development of TROR

Formal reanalysis and semantic reinterpretation are accompanied by a subset of correlated primitive changes. The primitive changes that apply to TROR are represented by the shaded cells in Table 43.

In the development of TROR, the level of phonology/phonetics remains unaffected, as far as written data is concerned.<sup>92</sup> Both morphological compositionality and the morphosyntactic properties of TROR change on the way to discourse marker. The compositionality of TROR changes in the sense that it forms a unit (syntagm>lexeme), whenever it occurs as a discourse marker. In all other cases it is still a VP. The discourse marker TROR is restricted to first-person pronouns.

In case of the discourse marker TROR it can be observed that its syntactic autonomy and variability increase because it becomes a syntactically independent unit. In all other uses, TROR

<sup>92</sup> But see Dehé and Wichmann (2010) for an explanation of the structural status of *I think* in terms of prosodic patterns.

is integrated into the syntactic structure, and affects the word order of the clause. The semantic compositionality of TROR is affected because the predominant meaning of this phrase is no longer its original literal meaning (=the cognitive act). That is, its meaning cannot be straightforwardly derived from its subparts. As regards the semantic substance of TROR there is weakening of the original cognitive meaning ‘believe (in) X’, whereas epistemic and communicative meanings have been added.

The fact that TROR contains a first-person subject may be a reason to assume that this phrase is inherently subjective. However, describing a cognitive act like ‘thinking’ or ‘believing’ is not subjective in the sense of expressive subjective meanings. Recall that the development of epistemic meaning always involves subjectification. In case of TROR, this pertains to the semantic shift from ‘cognitive state’ to ‘assumption’. A further, intersubjective shift occurs when TROR is used to express communicative meanings. Hence, also in case of TROR there has been a shift from syntactic to speaking subject, which coincide in this case.

<b>ii. Primitive changes</b>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Pgzn</b>
<b>-phonology/phonetics</b>			
° loss of phonological/phonetic substance	(+)	(+)	(+)
<b>-morphology</b>			
° loss of morphological compositionality	+	(+)	(+)
° loss of morphosyntactic properties	-	+	(+)
<b>-syntax</b>			
° loss of syntactic variability	-	+	-
° loss of syntactic autonomy	-	+	-
<b>-semantics</b>			
° loss of semantic substance	-	+	+
° loss of semantic compositionality	+	(+)	(+)
<b>-discourse/pragmatics</b>			
° subjectification	(+)	(+)	+
° intersubjectification	(+)	(+)	+

**Table 43.** Primitive changes in the development of TROR.

### 5.5.3 Side effects in the development of TROR

The side effects of formal reanalysis and semantic reinterpretation and their accompanying primitive changes can be used as a diagnostic to identify potential cases of lexicalization, grammaticalization or pragmaticalization. As regards TROR, the side effects that are noticeable in its development are represented by the shaded cells Table 44.



<b>iii. Side effects of change</b>	<b>Lxn1</b>	<b>Gzn1</b>	<b>Pgzn</b>
-paradigmaticization	-	+	(+)
-obligatorification	-	(+)	-
-condensation	-	+	-
-layering/divergence/specialization/persistence	+	+	+
-productivity	-	+	+
-frequency	-	+	+
-typological generality	-	+	(+)

**Table 44.** Side effects in the development of TROR.

The development of TROR shows *paradigmaticization* in the sense that there is a whole class of mental state predicates of the type ‘first person pronoun + cognitive verb in simple present tense.’ It is a *productive* expression because it attracts other forms and variations to the paradigm of mental state predicates, e.g. I WOULD THINK, I DON’T THINK etc. (cf. Van Bogaert (2010) for a constructional taxonomy of *I think* and its variant forms). TROR is an expression of high *frequency*, which has equivalent forms in other Germanic languages. Phrases like TROR are not grammatically obligatory, but may be ‘communicatively obligatory’. Condensation does not apply because the discourse marker TROR has not been subject to structural scope reduction and stronger internal dependencies within the clause.

The different co-existing structural representations of TROR are instances of layering, divergence, specialization and persistence. These days, TROR shows a lot of synchronic variation both formally and semantically. The cognitive verb continued to exist along with the epistemic and discourse variants. TROR is the most prototypical member of the class of mental state predicates. The original meaning of ‘believe (in) X’ still shines through in its later semantic developments.

#### **5.5.4 Linguistic status and degrees of X-ization for TROR**

As defined in the present study, lexicalization results in a linguistic item with referential meaning, primary status, and which may convey the main point of linguistic message, grammaticalization leads to a linguistic item with relational meaning, secondary status, and which regulates grammatical structure, and pragmaticalization results in a linguistic item with communicative meaning, extra-propositional status, and which organizes discourse structure.

With regard to TROR, its structural status is hard to establish, and highly context dependent. The patterns that arise from Table 42-Table 44 show that the development TROR has characteristics of lexicalization I, primary grammaticalization and pragmaticalization. The discourse marker bears the essential hallmarks of pragmaticalization, for primary grammaticalization only the semantic criteria are met and with respect to lexicalization I only the formal requirements are fulfilled. Because there is no formal reanalysis from VP to epistemic adverb, TROR is not an instance of grammaticalization. Instead, for the CTP TROR it seems more appropriate to speak of a secondary lexical status (cf. Boye & Harder 2007; 2012).

The types of subjectification involved in the development of TROR are summarized in Table 45 below.

<b>(Inter)Subjectification</b>	<b>TROR</b>
<b>I. subjectification</b> [speaker perspective, attitude and judgment]	+
- ideational level [meta-linguistic meanings] epistemic meaning, intention	
<b>III. intersubjectification</b> [interaction with interlocutor]	+
-interpersonal level communicative meanings (mitigation)	

**Table 45.** Types of subjectification in the development of TROR.

In case of the CTP TROR, only subjectification has occurred. There has not been a categorical reanalysis from VP into an epistemic adverb, hence it is a subjectified lexical CTP that has secondary status. The discourse marker is a (inter)subjectified syntactically independent unit (parenthetical), and hence a case of pragmaticalization.

## 5.6 Summary, discussion and conclusions

It has often been observed that mental state predicates like *I think* have an adverbial distribution, i.e. that these kind of phrases have been reanalyzed as epistemic parentheticals. The results of this case study show that Mainland Scandinavian TROR may indeed have adverbial-like properties, but only when it is a discourse marker. However, these uses constitute a minority in the samples. A vast majority of the corpus data are clauses in which TROR has subordinating properties because it essentially is a verbal predicate (cf. also MAYBE and MON). This goes to show that Mainland Scandinavian TROR is in full development, which is reflected by its distributional variation, both formally and semantically. Boye & Harder (2007:590-1) arrive at a similar conclusion in (173):

- (173) [T]he CTP at hand [i.e. in Danish] may not have evolved a grammatical variant, and even if it has, the syntactic and prosodic cues may be absent. [...] Cross-linguistically, there is substantial evidence for the movement from A to C [illustrated in (155)]. In some languages, CTP's and CTP clauses have not gone beyond the lexical-but-secondary position B. [...] In other cases, however, CTPs and CTP clauses resemble English *think* in *I think* in that they have reached position C. Often, the grammatical items that result from the development illustrated in (155) no longer have the same morphosyntactic form as a lexical clause [e.g. the epistemic adverb 'maybe'].

To conclude, in its epistemic sense TROR is a subjectified complement taking predicate; in its communicative function it developed into a discourse marker. When compared to English *I think*, the Mainland Scandinavian variants have not all reached stage C. The discourse marker

has, but this is only in its incipient stages (parenthetical TROR is a variable form, which may be ambiguous with complementing TROR). The CTP variant has not progressed beyond stage B.

## Chapter 6.

### 6. Discussion and conclusions

Many instances of change that have been analysed in the framework of grammaticalization studies defy strict categorization, either because they share properties of grammaticalization and lexicalization, or because they share some properties of grammaticalization, but not all of them. Some changes may even be ambiguous between lexicalization, grammaticalization and pragmaticalization (cf. Section 2.2)

In this study, three different case studies from the domain of epistemic modality have been discussed, namely the modal auxiliaries *MUST/MAY*, the epistemic adverb *MAYBE*, the interrogative/epistemic adverb *MON*, and the mental state predicate *TROR*. These cases are all challenging traditional grammaticalization analyses, because they are at the interface of (most current definitions of) grammaticalization, lexicalization and/or pragmaticalization, but they also differ from one another, so that it would not make sense to assign a special status the development of epistemic expressions. The introduction of new labels for problematic cases, resulting in a plethora of *izations*, only adds to the conceptual fuzziness around grammaticalization and related concepts. On the other hand, stretching the notion of ‘grammar’, and hence of ‘grammaticalization’ so as to include hybrid cases obscures substantial differences between different types of change, which reduces grammaticalization to a heterogeneous category with little descriptive power.

Section 6.1 presents a unified account of grammaticalization and related phenomena centered on the notion of a composite change. A composite change is composed of formal reanalysis and semantic reinterpretation, accompanying primitive changes at different linguistic levels, and the concomitants of these (micro-)changes. These changes cluster with respect to their converging properties (=interfaces), but become isolated in case they are diverging or unique properties.

A clustering analysis (cf. Norde & Beijering 2012) will be employed to detect the prototypical features of the investigated epistemic expressions in the present study, i.e. what are their converging (=shared features) and diverging properties(=unique properties). In addition, the case studies are compared to prototypical instances of *X-ization* from the literature to illustrate to what extent they deviate from standard cases of lexicalization, grammaticalization or pragmaticalization.

This final chapter presents a comparison of the investigated epistemic expressions and comments on the overall results of the present study within a unified account of language change. Section 6.2 is concerned with the main conclusions of this study and Section 6.3 provides an outlook to further research and future work.

## **6.1 A clustering approach to epistemic expressions at the lexicalization-grammaticalization-pragmaticalization interface**

The main idea, as outlined in the previous chapters, is that lexicalization, grammaticalization and pragmaticalization are composite changes that consist of i) formal reanalysis and semantic reinterpretation, ii) primitive changes at different linguistic levels, and iii) the side effects of i and/or ii. The converging and diverging properties of lexicalization, grammaticalization and pragmaticalization enable one to identify the unique and distinctive properties (diverging properties) of the different X-izations, as well as their shared properties and interface areas (converging properties). Throughout this study, the scheme has been applied to one case study at a time. In order to compare several case studies at the same time, the scheme has to be extended and adapted.

The first step in analyzing composite changes is to select the type of formal reanalysis and semantic reinterpretation. The first part of the scheme functions as a filter because it identifies the distinctive mechanisms involved in a change. The essential mechanisms in primary and secondary grammaticalization, lexicalization I and II, and pragmaticalization are highlighted, and as such it provides an indication of the most likely X-ization for a given development.

The next step in examining composite changes is to reduce them to their primitive changes (cf. Norde 2009:36). Primitive changes apply at different linguistic levels, i.e. phonology/phonetics, morphology, syntax, semantic and discourse/pragmatics. They are defined in terms of expansion and reduction (cf. Traugott 2010a) at different linguistic levels, and in principle, independent of the type of X-ization. Primitive changes may accompany reanalysis and reinterpretation, for example, loss of morphosyntactic properties comes along with a categorical reanalysis from major to minor category.

The final step in exploring composite changes is to look at the concomitants of formal reanalysis, semantic reinterpretation and the various correlated primitive changes at different linguistic levels. The side effects only tell that a change has occurred, but they remain indecisive on what kind of change. In general, every category shift results in a newly developed form besides the original form, which implies divergence, layering, specialization and persistence. The concomitants of change only serve a descriptive purpose. They are not useful in determining types of X-ization.

All in all, a full analysis of composite changes consists of an indication of the type of X-ization based on the type(s) of reanalysis and reinterpretation, clusters of reductive and expansive primitive changes, as well as the concomitants of these changes. All these changes collectively identify a change as either a case of grammaticalization, lexicalization or pragmaticalization, or, as a hybrid/tripartite case at the interface of these different types of language change.

The inclusion of all possibilities in one scheme, both the plausible and implausible, as well as distinctive and optional properties of a certain X-ization, enables one to look at a development in an unbiased way. It allows one to exclude certain scenarios, but at the same time one has to keep in mind that there may be other available options than the one which is most plausible at first sight. For instance, when asking whether a certain development counts as an instance of grammaticalization, one is inclined to look for properties of grammaticalization only,

and overlook other scenarios. As a result, developments may be classified as (non-prototypical) instances of grammaticalization or dismissed as an instance of grammaticalization in case the development does not match (all of) the chosen grammaticalization criteria. By contrast, this unified scheme of language change is capable of identifying prototypical instances, as well as marginal cases, of a certain X-ization, and it identifies interface areas between different types of languages change.

In order to see if, and to what extent, the development of the different epistemic expressions coincide, they are compared to one another, and to ‘standard cases’ of primary grammaticalization (*to be going to*, e.g. Fischer & Rosenbach 2000), secondary grammaticalization (definiteness marking from word (clitic) to inflectional suffix in North Germanic, (Proto-Nordic *hús it* ‘house that(n)’ > Old Norse *húsit* ‘house-def(n)’, cf. Enger 2012), i.e. suffixed article in present-day Scandinavian), lexicalization I, fusion, (*lord*, e.g. Brinton & Traugott 2005), lexicalization II, separation, (*ism*, e.g. Antilla 1989), and pragmaticalization (*you know*, Aijmer 1997), as described in the literature and in Chapter 2.

These prototypical instances are only included for comparison. They show to what extent these hybrid expressions deviate from prototypical cases of the different X-izations. The symbols that represent the standard cases of X-izations and the case studies are presented in (174) below.

(174)	<b>symbol</b>	<b>‘standard case’ of X-ization</b>
	✱	<i>to be going to</i> (primary grammaticalization)
	✱	<i>hus.et</i> suffixed article (secondary grammaticalization)
	☪	<i>lord</i> (lexicalization I (fusion))
	▲	<i>ism</i> (lexicalization II (separation))
	♣	<i>you know</i> (pragmaticalization)
	<b>symbol</b>	<b>case study</b>
	◆	<b>MUST/MAY</b>
	●	<b>MON</b>
	■	<b>MAYBE</b>
	♠	<b>TROR</b> (complement taking predicate: ♠ <sub>C</sub> , discourse marker: ♠ <sub>DM</sub> )

Ideally, the standard cases should not be characterized by properties that are unique of changes other than the one to which they are assigned. The case studies, which are concerned with hybrid items, have the potential of being characterized by the unique properties of more than one X-ization.

### 6.1.1 Mechanisms in the development of MUST/MAY, MON, MAYBE and TROR

The first part of the scheme identifies the type of reanalysis and reinterpretation involved in a change. With regard to the case studies and prototypical instances of the different X-izations, the scheme looks as in Table 46 below.

<i>i. Mechanisms of change</i>	X-ization	case study /standard case of X-ization
<b>hierarchical reanalysis</b>		
propositional > extra-propositional status	Pgzn	♣ ♠ <sub>DM</sub>
<b>categorical reanalysis</b>		
major > minor category	Gzn1	* ♦ ■
minor > minor category	Gzn2	♣ ●
<b>constituent internal reanalysis*</b>		
syntagm/complex lexeme > (simple) lexeme	Lxn1 (Gzn1) (Gzn2) (Pgzn)	☐ * ■ ♣ ♣ ♠ <sub>DM</sub>
bound morpheme > semi-independent word	Lxn2	▲
<b>reinterpretation (metaphor/metonymy)</b>		
referential > referential meaning	Lxn1	☐
referential > relational meaning	Gzn1	* ♦ ■
relational > relational meaning	Gzn2	♣ ♦ ●
referential/relational > referential meaning	Lxn2	▲
referential/relational > communicative meaning	Pgzn	♣ ● ■ ♠

**Table 46.** Mechanisms of change in standard cases of X-ization and the case studies.

The investigated epistemic expressions in the present study have all undergone formal reanalysis and semantic reinterpretation. All items shifted category, with the exception of TROR and Swedish *måste* which was borrowed from Middle Low German as a modal form. Only MAYBE is the result of a constituent internal reanalysis in that it is a univerbation of the modal form MAY and the main verb HAPPEN. In case TROR is a discourse marker, it is reanalyzed as an extra-propositional unit, which has not (yet) been fixed and fused.

Semantically, there are shifts from referential meaning to relational meanings (primary grammaticalization) for MAYBE and the modals MAY/MUST. The modals also witness consecutive shifts from relational to other relational meanings. The adverbial epistemic expressions have also developed speech-act meanings, which are reflected by the rise of (inter)subjective meanings.

### 6.1.2 Primitive changes in the development of MUST/MAY, MON, MAYBE and TROR

The second part of the scheme specifies the reductive and expansive primitive changes at different linguistic levels that may accompany lexicalization, grammaticalization and

pragmaticalization. The middle position  $\pm$  indicates that there is no change with respect to a certain linguistic level, i.e. the change is not applicable or the parameter is unaffected.

With respect to syntax there are two parameters, variability and autonomy. Variability denotes the flexibility of a linguistic item, i.e. the number of syntactic slots that an element may occupy. Autonomy is a related concept that concerns the degree of syntactic integration for a given item.

The two parameters of morphological change are morphosyntactic properties and morphological compositionality. The former relates to definiteness, inflectional properties such as tense, case and number, and subcategorization features. For instance, when an item shifts category status it will lose the characteristics typical of its original category. Morphological compositionality relates to the form of a linguistic item, it concerns the degree of formal transparency or opacity of compositional linguistic items. For example, constituent internal reanalysis may result in unanalyzable forms (e.g. OE *hlaf* ‘loaf’ + *weard* ‘guardian’ > *lord*).

Changes at the level of phonology and phonetics apply to the phonological substance of an item, and are closely linked to morphosyntactic properties. There may be loss of segments, prosodic shifts, or the realization of an item may be subject to weakening or strengthening. This may all result in reduced word length (e.g. *gonna*) but also more prominent items (e.g. *ism*).

As regards semantics, the two parameters are semantic substance and semantic compositionality. In this scheme semantic substance concerns referential meaning. As referential meanings fade, relational meanings become more salient. Semantic compositionality relates to the extent to which the meaning of an expression can be derived or constructed from its subparts. It is about the degree of semantic transparency or opacity of compositional forms.

Discourse and pragmatics relate to the communicative aspects of an expression. They pertain to the degree of speaker-perspective and speaker-addressee interaction. There are different types of subjectification, which apply to the ideational, textual and interpersonal level. The rise of (inter)subjective meanings is indicative of change with respect to discourse/pragmatics.

As regards the case studies and prototypical instances of the different X-izations, the scheme for primitive changes looks as in Table 47 below. For ease of comparison, the symbols that represent the case studies and standard cases are repeated in (175) below.

(175)	<i>symbol</i>	<i>x-ization</i>		<i>symbol</i>	<i>case study</i>
	✱	<i>to be going to</i>		◆	<b>MAY/MUST</b>
	✱	<i>hus.et</i>		●	<b>MON</b>
	☪	<i>lord</i>		■	<b>MAYBE</b>
	▲	<i>ism</i>		♠CTP, ♠DM	<b>TROR</b>
	♣	<i>you know</i>			



<i>ii. Primitive changes</i>				
linguistic level	parameter	reduction (-)	±	expansion (+)
<b>syntax</b>	syntactic variability	* ♦	☐ <sub>CTP</sub>	▲ ♦ ● ■ ☐ <sub>DM</sub>
	syntactic autonomy	* ♦	☐ <sub>CTP</sub>	▲ ♦ ● ■ ☐ <sub>DM</sub>
<b>morphology</b>	morphosyntactic properties	* ♦ ■ ☐ <sub>DM</sub>	☐ <sub>CTP</sub> ●	▲
	morphological compositionality*	* ☐ <sub>DM</sub>	● ▲ ☐ <sub>CTP</sub> ● ♦	
<b>phonology/phonetics</b>	phonological substance	* ♦ ☐	● ☐ ♦	▲
<b>semantics</b>	semantic substance	* ♦ ♦ ■ ☐ ●	☐	▲
	semantic compositionality*	* ☐ ♦ ☐	● ▲ ● ♦	
<b>discourse/pragmatics</b>	subjectification		☐	▲ ♦ * ♦ ● ■ ☐
	intersubjectification		* ☐ ▲	♦ ● ■ ☐ ♦

**Table 47.** Primitive changes in standard cases of X-ization and the case studies.

As regards the reductive and expansive primitive changes that may accompany formal reanalysis and semantic reinterpretation, we see that there is expansion at the discourse/pragmatics level for all epistemic expressions, and expansion at the syntactic level for the adverbial expressions only. There is reduction with respect to semantic substance for all epistemic expressions. The reductive and expansive primitive changes will be separately discussed and compared for the case studies, in 6.2.1-6.2.4 below.

### 6.1.2.1 Syntax

The reanalysis from verb phrase (MAYBE and TROR) or auxiliary (MON) to sentence/speech act adverb necessarily results in increased *syntactic variability* because adverbs are more flexible elements by nature. In case of MAYBE and MON this means that they are gradually acquiring sentence adverbial distributions. The discourse marker TROR is an instance of syntactic externalization, which means that the item is no longer part of the proposition, but an additional comment towards the proposition, formalized as an adjunct or extra-sentential element. For modal auxiliaries, on the other hand, the reanalysis from main verb to auxiliary inevitably leads to a decrease in syntactic variability as auxiliaries are syntactically more restricted than main verbs and obtain fixed syntactic slots.

As regards *syntactic autonomy*, which relates to the degree of syntactic integration and dependencies in the linguistic system, it can be observed that verb phrases and auxiliaries are more tightly integrated into the syntactic structure than various types of adverbs and discourse

markers. The modal auxiliaries MAY/MUST lose in syntactic autonomy as they become dependent on the main verb of the clause. With respect to MAYBE, which may occur in V2 and non-V2 structures, the instances of MAYBE in non-V2 structures are more integrated than the occurrences in V2 structures. In non-V2 clauses MAYBE has so-called ‘subordinating properties’, i.e. MAYBE is followed by subordinate word order. In V2 clauses MAYBE takes sentence adverbial positions, or may be an adjunct. The interrogative adverb MON has retained the subordinating properties of its verbal origin in Norwegian, in Danish it may occasionally be used as a sentence adverb, and in Swedish it functions solely as a sentence adverb. As a discourse marker, TROR does not have influence on syntactic structures, as it is no longer part of the proposition, but a comment towards the proposition.

### 6.1.2.1 Morphology

The loss of *morphosyntactic properties* follows from categorical reanalysis, in case of a shift from major to minor category. All epistemic expressions in this study ultimately derive from full verbs or verb phrases. When transforming to another word class, i.e. auxiliary or adverb, they lose the inflectional properties of verbal paradigms. For example, once MAYBE is a univerted adverb, the inflectional properties and variability of the verb phrase have been lost. The same applies to MON and the synonymous adverb MONTRO. The modal auxiliaries MAY/MUST do not have a number of features characteristic of verbal paradigms, such as passive, imperative and participle forms. Swedish *må*, *måtte* also lost the infinitival marker and the ability to inflect for tense. Swedish *måste* was borrowed as ‘ready-made’ modal with a defective inflectional paradigm. Danish and Norwegian *måtte* still have infinitival markers and do inflect for tense. For the mental state predicate TROR, the prototypical form consists of a first person singular subject and the verb in present tense.

The related parameter of *morphological compositionality* refers to the degree of transparency or analyzability of compositional items and expressions. Only MAYBE and the discourse marker TROR are compositional forms. The origin of these items is transparent. For MAYBE, the modal form MAY and the main verb HAPPEN are identifiable, for TROR the pronoun I and cognitive verb ‘think, believe’ are recognizable. With respect to TROR, which may convey referential, epistemic or communicative meanings, there is not (yet) a distinct morphosyntactic status for each of its different variants (MC, CTP and DM).

### 6.1.2.3 Phonology/phonetics

As observed with respect to morphosyntactic properties, the form as well as the *phonological substance* of the items has not changed significantly. Only MAYBE is an instance of constituent internal reanalysis, which in the process of univertation, may have lost the referential subject IT and the complementizer THAT, in case it developed out of a matrix clause (Wessén 1967). In Norwegian and Swedish there has also been a prosodic shift in that verb phrase accentuation changed into adverb accentuation, i.e. previously there was stress on both MAY and BE, now only MAY takes stress. This prosodic shift did not occur in Danish. All other epistemic phenomena

have not witnessed any audible changes. In so far written data can tell there are no visible instances of loss or reduction of phonetic/phonological segments in the written data.

#### 6.1.2.4 Semantics

The *semantic substance* for all epistemic expressions is reduced in the sense that they no longer denote the original literal meaning. Their meanings have become more abstract, general and functional. The epistemic adverb MAYBE expresses epistemic possibility rather than literally stating that something ‘can happen’. The modal meaning of MON was reanalyzed into an interrogative marker expressing various dimensions of doubt on behalf of the speaker. The modal auxiliaries developed a wide range of modal and postmodal meanings, all of which are more abstract than the original lexical meaning ‘to have the power/strength’ or ‘to have the opportunity’. In case of TROR there has been a change from act of cognition, to mode of knowing, to epistemic possibility (cf. Brinton 1996), and opinion. The semantic development of the epistemic phenomena under investigation is characterized by contiguous, metonymic changes. However, through the course of time these associations fade away, which makes items non-transparent if one is not acquainted with its history.

*Semantic compositionality* relates to the extent to which the meaning of an expression can be derived or constructed from its subparts. Both the subparts and the entire expressions no longer express their literal meaning. In case of *kanske*, *kanskje* and *måske* the meaning of *sk(j)e* ‘happen’ got lost and replaced by the more general ‘be’. As regards TROR, the qualificational meanings have taken over the literal meaning of ‘being in cognitive state X.’ In short, the meaning of these epistemic expressions is not the sum of the (literal) meanings of their subparts.

#### 6.1.2.5 Discourse/pragmatics

Discourse and pragmatics relate to the communicative aspects of a linguistic expression. The development of epistemic meaning always involves increased speaker-perspective (subjectification), which is reflected by a shift from syntactic to speaking subject. The adverbs MAYBE and MON developed into epistemic markers as well as speech-act adverbs expressing the speaker’s point of view with respect to possibility. The modal auxiliaries came to express personal evaluations of possibilities and probabilities. They also have directive meanings, such as permission and obligation. Especially Swedish *må* and *måtte* occur in a number of set phrases and are frequently used in dialogue and exclamations. The discourse marker TROR conveys the speaker’s assumption or intention.

Different types of subjectification apply to the epistemic expressions under investigation. This is primarily related to the linguistic layers to which they apply, as summarized in Table 48.

	<b>(Inter)Subjectification</b>	
	<b>I. subjectification</b> [speaker perspective, attitude and judgment]	
* <i>to be going to</i> * <i>hus.et</i> ◡ <i>lord</i> ▲ <i>ism</i> ♣ <i>you know</i> ----- ♦ MAY/MUST ● MON ■ MAYBE ♠ TROR	- ideational level [connotation: amelioration and pejoration] e.g. <i>boor</i> ‘farmer’ > <i>boor</i> ‘crude person’  - ideational level [meta-linguistic meanings] e.g. <i>observe</i> ‘perceive (that)’ > <i>observe</i> ‘state that’  -textual level [text-construction] e.g. connective <i>while</i> [ <i>þa hwile þe</i> ‘the time that’ > <i>while</i> ‘during’]  -textual level [meta-linguistic meanings] e.g. modal auxiliaries <i>can, may, must</i> etc.  -interpersonal level e.g. discourse markers <i>I mean, I think</i> etc.	▲  ♠CTP  *●  *♦●■  ♣♠DM
	<b>II. intersubjectification</b> [interaction with interlocutor]	
	-ideational level e.g. conversational routines, <i>thanks, goodbye, please</i> etc.  -textual level e.g. modal particles <i>vel, jo, nok, da</i> etc. in Norwegian  -interpersonal level e.g. discourse markers <i>you know/y’know, look-forms, well</i> etc.	  ♦●■  ♣♠DM

**Table 48.** Types of subjectification in standard cases of X-ization and the case studies.

The rise of MUST/MAY is an instance of grammaticalization accompanied by subjectification (epistemic meanings) and (inter)subjectification (concessive and optative meanings) at the textual level. The development of MON involves shifts within the textual level, which is a case of secondary grammaticalization accompanied by (inter)subjectification. The development of MAYBE is a shift from the ideational to the textual level and a case of grammaticalization accompanied by (inter)subjectification. In case of the CTP TROR, only subjectification has occurred. There has not been a categorical reanalysis from VP into an epistemic adverb. Hence it is a subjectified, secondary lexical CTP. The discourse marker TROR is an (inter)subjectified syntactically independent unit, and hence a case of pragmaticalization.

### 6.1.3 Side effects in the development of MUST/MAY, MON, MAYBE and TROR

The third part of the scheme lists possible side effects or concomitants of a change. Most of these heuristics have been mentioned in relation to grammaticalization (Hopper 1991; Lehmann 1995) or the lexicalization-grammaticalization interface (Brinton & Traugott 2005). Likewise, the absence of these side effects suggests that the change in question does not qualify as a prototypical instance of grammaticalization. The side effects indicate that reanalysis, reinterpretation and accompanying primitive changes have occurred, but they cannot be used to uniquely identify instances of a certain X-ization, see Table 49.

<i>symbol</i>		<i>iii. Side effects of change</i>	<i>X-ization</i>	<i>case study/standard case</i>
* <i>to be going to</i> * <i>hus. et</i> ☪ <i>lord</i> ▲ <i>ism</i> ♣ <i>you know</i> ----- ♦ MAY/MUST ● MON ■ MAYBE ♠ TROR -----		paradigmaticization	Gzn1 Gzn2 (Pgzn)	* ♦ ■ ♠
		obligatorification	Gzn2 (Gzn1)	*
		condensation	Gzn1 Gzn2	* * ♦
		layering divergence specialization persistence	Lxn2 Gzn1 Gzn2 Pgzn	* * ▲ ♣ ♦ ● ■ ♠
		productivity	Lxn2 Gzn1 Gzn2 Pgzn	* * ▲ ♣ ♦ ■ ♠
		frequency	Lxn2 Gzn1 Gzn2 Pgzn	* * ▲ ♣ ■ ♦ ♠
		typological generality	Gzn1 (Lxn2) (Gzn2) (Pgzn)	* ▲ ♣ ■ ♦ ♠

**Table 49.** Side effects in standard cases of X-ization and the case studies.

All epistemic expressions in the present study are the result of formal reanalysis and semantic reinterpretation. Because of this, there is a state of synchronic variation, or layering, of different form-meaning pairs. In the development of Swedish *må* there has been a lexical and grammatical split, the former resulting in the lexical verb *att må* ‘to feel’ and the latter relates to the rise of the

two different modal forms *må* and *måtte*.

The rise of epistemic expressions does not result in obligatory grammatical items, such as inflections, case, or agreement, because epistemicity can be expressed in various ways, none of which is obligatory. Epistemic judgments may, however, be communicatively required. Only the modal auxiliaries become more to tightly integrated and dependent within the syntactic structure. Increased productivity and paradigmaticization concern context-expansion (type frequency) as well as becoming a member of another paradigm. It is closely related to the frequency with which an item is encountered (token frequency). Swedish *må* and *måtte*, and MON in general, have reached the end stage of their development and have become highly specialized, context-dependent items, which reduces their frequency and productivity. The opposite is true for MAYBE, TROR, Danish *måtte*, and Norwegian *måtte*. These expressions are part of different (sub)paradigms and they are of high frequency. The rise of MON seems to be unique to the Mainland Scandinavian languages, whereas the development of modal auxiliaries like MUST/MAY, epistemic adverbs of the MAYBE-type, and parenthetical phrases like TROR do have typological correlates in other languages.

## 6.2 Conclusions

The general picture that arises when visualizing the changes involved in the case studies and the prototypical instances of the X-izations, is that they primarily relate to the degree of the overall autonomy of an element within the linguistic system (cf. Lehmann 1995). Grammatical items typically lose autonomy and substance on all linguistic levels because movement towards core grammar is accompanied by tighter integration and dependencies within the linguistic system. Lexical items resulting from fusion change with respect to compositionality (semantics + morphology), but there is no change of category leading to altered dependencies within the system. The autonomy of fused lexical items remains unaffected as they remain subject to the general rules of grammar and word combining. Lexical items resulting from separation are on their way of becoming autonomous words. Communicative items, on the other hand, are not subject to the rules of grammar but operate at the level of discourse. They witness an increase in autonomy, in the sense that they become syntactically independent units, which function as additional comments towards the proposition.

In short, lexical items are autonomous, independent items at the ideational level, grammatical items are ancillary, dependent items at the textual level, and discourse markers are autonomous, extra-propositional items at the interpersonal level. These observations, as well as the scheme of converging and diverging properties of the different types of language change show that pragmaticalization is genuinely different from lexicalization and grammaticalization, and as such should be defined in its own right.

In order to distinguish between lexicalization, grammaticalization and pragmaticalization it is important to examine the status of an item with respect to the proposition and to distinguish different components of language: lexicon (ideational level, i.e. words and expressions with referential meaning), grammar (textual level, i.e. grammatical functions structural dependencies)

and discourse (interpersonal level, i.e. communicative comments towards the proposition). Lexicalization (both fusion and separation) operates within the ideational level, grammaticalization applies to the textual level (a shift from the ideational to textual level, or within the textual level), and pragmaticalization operates at the interpersonal level (a shift from ideational or textual level to interpersonal level). Note that in case of an X-ization the entire form-meaning pair shifts from one level to another. Subjectification applies at the different layers and may accompany instances of X-ization, but this need not be the case. For example, the Dutch evaluative adjective *dom* (cf. de Smet & Verstraete 2006, Section 2.1.4) is not an instance of lexicalization but it is nonetheless subject to subjectification. Likewise, epistemic TROR as an (lexical) CTP is subjectified but not (yet) lexicalized/grammaticalized. The shifts between the different layers are, amongst others, reflected by the existence of hybrid/tripartite items at the lexicalization-grammaticalization-pragmaticalization interface.

The case studies in the current investigation, but also the many hybrid cases mentioned in Chapter 2, defy strict categorization. Categorization is a useful theoretical construct but it is generally not applicable to actual instances of change. In this study, the idea of categorizing developments into predefined types of change is given up. This does not mean that the concept of grammaticalization as such is not useful (see Section 1.3), but that it cannot be defined in terms of an (arbitrary) set of formal and semantic primitive changes. Instead, a clustering approach to grammaticalization and related phenomena has been applied. The advantage of clustering to categorization is that a clustering approach can deal with the gradient nature of primitive changes and lexicalization, grammaticalization and pragmaticalization in general, and captures hybrid cases and interface areas, i.e. it indicates where different types of language change coincide. In addition, a cluster approach enables one to identify both prototypical and marginal instances of certain types of language change.

The clusters may coincide with the sum of changes that traditionally represent ‘lexicalization’, ‘grammaticalization’ or ‘pragmaticalization’, but they may also pattern in alternative ways. This is what happens with respect to the case studies presented in Chapter 3-5 because these are all instances of developments at the lexicalization-grammaticalization-pragmaticalization interface. All in all, the investigated epistemic phenomena have much in common, but they also differ considerably with respect to syntactic properties. These dissimilarities may largely be due to differences in word class as adverbs are more ‘free’ items than auxiliaries. The common denominator in the development of these epistemic expressions is the expression of speaker perspective, attitude and judgment, i.e. subjectification.

### 6.3 Outlook and suggestions for further research

Throughout this study a number of topics that go beyond the scope of the present study were mentioned in passing. One of these is the role of prosody in determining the status of linguistic items (i.e. lexical, grammatical or communicative), which importance has repeatedly been mentioned in the literature (e.g. Dehé & Wichmann 2010). That is, in a chain of developments, is it the case that the full referential form is also prosodically most prominent? Related to this point,

is the study of spoken language. This study is based on data from written corpora of, amongst others, newspaper texts. It would be interesting to see if the results of this corpus study also extend to spoken data, or other text genres.

From a typological point of view, the inclusion of languages other than the Germanic languages, would cast more light on the universality of the identified tendencies here, and in the literature. From a comparative perspective, parallel corpora would provide information about the subtle differences in languages with respect to epistemic representations. The focus of this investigation has been on three case studies that were concerned with adverbs, modals and discourse markers. In order to get a full picture of the structure and complexity of epistemic modality, it would be interesting to see how (the development of) other epistemic expressions, such as adjectives, nouns, mood, particles, modal collocations and modifiers, cluster with respect to mechanisms of change, accompanying primitive changes and concomitants of change.

The cluster approach has now been applied to epistemic expressions at the lexicalization-grammaticalization-pragmaticalization interface, but may also be used to compare other types of ‘unclassifiable changes’, as well as non-controversial instances of the different X-izations. The scheme can easily be extended so as to include degrammaticalization phenomena as well, so that it captures all types of language change in one overview. In order to get an idea of the changes that typically go along with a certain X-ization, domain (e.g. modality), linguistic category (e.g. adverbs), etc., larger data sets (=as many as possible attested instances) are needed to establish robust tendencies in their overall development. Thus, a clustering approach enables one to make larger comparisons and to detect general tendencies in languages change.



## Appendices

		<b>Danish <i>må</i></b>	<b>Norwegian <i>må</i></b>	<b>Swedish <i>må</i></b>	<b>Swedish <i>måste</i></b>
<b>deontic</b>	count	417	463	121	432
	percentage	83,4%	92,6%	12,1%	86,4%
<b>epistemic</b>	count	78	37	22	55
	percentage	15,6%	7,4%	2,2%	11,0%
<b>optative</b>	count	3	0	83	0
	percentage	0,6%	0%	8,3%	0%
<b>concessive</b>	count	0	0	537	1
	percentage	0%	0%	53,7%	0,2%
<b>excluded</b>	count	2	0	237	12
	percentage	0,4%	0%	23,7%	2,4%
<b>total</b>	count	500	500	1000	500
	percentage	100,0%	100,0%	100,0%	100,0%

**Appendix 1.** Cross table of the frequencies for the different semantic categories of *må* (present tense) and *måste* in the Danish, Norwegian and Swedish random samples.

		<b>Danish <i>måtte</i></b>	<b>Norwegian <i>måtte</i></b>	<b>Swedish <i>måtte</i></b>	<b>Swedish <i>måste</i></b>
<b>deontic</b>	count	421	421	11	432
	percentage	83,9%	84,2%	8,2%	86,4%
<b>epistemic</b>	count	42	15	51	55
	percentage	8,6%	3,0%	38,1%	11,0%
<b>optative</b>	count	2	3	68	0
	percentage	0,4%	0,6%	50,7%	0%
<b>concessive</b>	count	5	0	4	1
	percentage	1,0%	0%	3,0%	0,2%
<b>eventuality</b>	count	22	19	0	0
	percentage	4,5%	3,8%	0%	0%
<b>excluded</b>	count	8	42	0	12
	percentage	1,6%	8,4%	0%	2,4%
<b>total</b>	count	500	500	134	500
	percentage	100,0%	100,0%	100,0%	100,0%

**Appendix 2.** Cross table of the frequencies for the different semantic categories of *måtte* (past tense) and *måste* in the Danish, Norwegian and Swedish random samples.

		<b>Danish <i>mon</i></b>	<b>Norwegian <i>mon</i></b>	<b>Swedish <i>månne</i></b>
<b>question particle</b>	count	78	75	0
	percentage	78 %	87.2 %	0 %
<b>concessive meaning</b>	count	3	0	0
	percentage	3 %	0 %	0 %
<b><i>mon det</i> (set phrase)</b>	count	0	11	0
	percentage	0 %	12.8 %	0 %
<b>epistemic adverb</b>	count	13	0	78
	percentage	13 %	0 %	100 %
<b><i>mon ikke</i> (set phrase)</b>	count	6	0	0
	percentage	6 %	0 %	0 %
<b>total</b>	count	100	86	78
	percentage	100 %	100 %	100 %

**Appendix 3.** Cross table of the frequencies for the different meanings and uses of MON in the Danish, Norwegian and Swedish random samples.

		<b>Danish <i>måske</i></b>	<b>Norwegian <i>kanskje</i></b>	<b>Swedish <i>kanske</i></b>
<b>SF</b>	count	177	342	114
	percentage	33.5 %	43.6%	22.6%
<b>KF</b>	count	225	237	184
	percentage	42.5 %	30.2%	36.4%
<b>OF</b>	count	21	23	10
	percentage	4 %	2.9 %	2 %
<b>AF</b>	count	95	94	51
	percentage	18 %	12 %	10.1 %
<b>SK</b>	count	0	0	74
	percentage	0 %	0 %	14.7 %
<b>OK</b>	count	0	0	5
	percentage	0 %	0 %	1 %
<b>AK</b>	count	2	0	19
	percentage	0.4 %	0 %	3.8 %
<b>KS</b>	count	9	88	48
	percentage	1.7 %	11.2 %	9.5 %
<b>total</b>	count	529	784	505
	percentage	100 %	100 %	100 %

**Appendix 4.** Cross table of the frequencies for the different syntactic structures with MAYBE in the Danish, Norwegian and Swedish random samples.

meaning		Danish <i>tror</i>	Norwegian <i>tror</i>	Swedish <i>tror</i>
non-qualificational	count	171	83	137
	% within language	17.1%	8.3%	13.7%
qualificational	count	829	917	863
	% within language	82.9%	91.7%	86.3%
total	count	1000	1000	1000
	% within language	100%	100%	100%

**Appendix 5.** Qualificational vs non-qualificational meanings of TROR.

person		Danish <i>tror</i>	Norwegian <i>tror</i>	Swedish <i>tror</i>
1	count	585	392	387
	% within language	70.7 %	45.0 %	41.7 %
2	count	48	25	44
	% within language	5.8 %	2.9 %	4.7 %
3	count	194	455	498
	% within language	23.5 %	52.2 %	53.6 %
total	count	827	872	929
	% within language	100 %	100 %	100%

**Appendix 6.** The distribution of first, second and third person pronouns with TROR.

person	that		Danish <i>tror</i>	Norwegian <i>tror</i>	Swedish <i>tror</i>
1	+that	count	177	32	205
		% within language	32.4%	9.4%	61.9%
1	-that	count	370	307	126
		% within language	67.6%	90.6%	38.1%
total		count	547	339	331
		% within language	100 %	100 %	100 %
3	+that	count	67	112	273
		% within language	34.5%	24.6%	63.5%
3	-that	count	127	344	157
		% within language	65.5%	75.4%	36.5%
total		count	194	456	430
		% within language	100 %	100 %	100 %

**Appendix 7.** The ratio for *that*-deletion vs *that*-complementation for TROR.

<b>person</b>	<b>position</b>		<b>Danish <i>tror</i></b>	<b>Norwegian <i>tror</i></b>	<b>Swedish <i>tror</i></b>
<b>1</b>	<b>clause-initial</b>	count	287	168	215
		percentage	52.5 %	55.8 %	65.0 %
<b>1</b>	<b>clause-medial</b>	count	245	130	111
		percentage	44.8 %	43.2 %	33.5 %
<b>1</b>	<b>clause-final</b>	count	15	3	5
		percentage	2.7%	1.0 %	1.5 %
<b>total</b>		count	547	301	331
		percentage	100 %	100 %	100 %
<b>3</b>	<b>clause-initial</b>	count	77	269	223
		percentage	39.9 %	58.5 %	51.9%
<b>3</b>	<b>clause-medial</b>	count	112	147	139
		percentage	58.0 %	32.0 %	32.3 %
<b>3</b>	<b>clause-final</b>	count	4	44	68
		percentage	2.1 %	9.6 %	15.8 %
<b>total</b>		count	193	460	430
		percentage	100 %	100 %	100%

**Appendix 8.** Clausal positions for TROR with first en third person subjects.

## References

- Abraham, Werner. 1991. The grammaticization of the German modal particles. In Elizabeth Closs Traugott & Bernd Heine (eds) *Approaches to Grammaticalization*, volume II, 331-380. Amsterdam/Philadelphia: John Benjamins.
- Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
- Aijmer, Karin. 1997. *I think* - an English modal particle. In Swan, Toril & Olaf Jansen Westvik (eds) *Modality in Germanic languages. Historical and comparative perspectives*, 1-47. Berlin/New York: Mouton de Gruyter.
- Aijmer, Karin. 2009. *Seem* and evidentiality. *Functions of Language* 68, 63-88.
- Aijmer, Karin & Anne-Marie Simon-Vandenberg. 2009. Pragmatic markers. In Östman, Jan-Ola & Jef Verschueren (eds) *Handbook of Pragmatics*. Amsterdam/Philadelphia: John Benjamins.
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Andersen, Henning. 2001. *Actualization and the unidirectionality of change*. In Andersen, Henning (ed) *Actualization: Linguistic Change in Progress*, 225-248. Amsterdam/Philadelphia: John Benjamins.
- Andersson, Peter. 2007. *Modalitet och förändring. En studie av må och kunna i fornsvenska*. [Modality and change. A study of *må* and *kunna* in Old Swedish]. Göteborg: Göteborgs Universitet.
- Andréasson, Maia. 2002. *Kanske- en vilde i satsschemat*. [Kanske- a savage in the sentence scheme] (=Meddelanden från Institutionen för svenska språket 41). Göteborg: Institutionen för Svenska Språket.
- Antilla, Raimo. 1989 [1972]. *Historical and Comparative Linguistics*. Second Edition. Amsterdam/Philadelphia: John Benjamins.
- Athanasiadou, Angeliki, Costas Canakis & Bert Cornillie (eds). 2006. *Subjectification: Various Paths to Subjectivity*. Berlin/New York: Mouton de Gruyter.
- Autenrieth, T. 2002. *Heterosemie und Grammatikalisierung bei Modalpartikeln. Eine synchrone und diachrone Studie anhand von eben, halt, ech(er)t, einfach, schlicht und glatt*. [Heterosemy and grammaticalization in modal particles. A synchronic and diachronic study on the basis of *eben, halt ech(er)t, einfach, schlicht and glatt*.] Tübingen: Niemeyer.
- Bauer, Laurie. 1983. *English Word Formation*. Cambridge, UK: Cambridge University Press.
- Beeching, Kate. 2009. Procatalepsis and the Etymology of Hedging and Boosting Particles. In Hansen, Maj-Britt Mosegaard & Jacqueline Visconti (eds) *Current Trends in Diachronic Semantics and Pragmatics*, 81-105. Bingley: Emerald.
- Bellert, Irene. 1977. On semantic and distributional properties of sentential adverbs. *Linguistic Inquiry* 8, 337-351.
- Benveniste, Emile. 1966. De la subjectivité dans le langage [Subjectivity in language]. In Benveniste, Emile (ed) *Problèmes de Linguistique Générale*, volume I, 258-266. Paris: Gallimard.

- Birkmann, Thomas. 1987. *Präteritopräsentia: morphologische Entwicklungen einer Sonderklasse in den altgermanischen Sprachen*. [Preterite-presents: morphological developments of a special class in the Old Germanic languages] Tübingen: Niemeyer.
- Björkstam, Harald. 1919. *De modala hjälpverben i svenskan. Tör, lär, mon, må, måtte och vill*. [The modal auxiliaries in Swedish. Tör, lär, mon, må, måtte and vill] Lund: Håkan Ohlssons Boktryckeri.
- Blank, Andreas. 2001. Pathways of lexicalization. In Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds) *Language Typology and Language Universals*, volume II, 1596-1608. Berlin/New York: Walter de Gruyter.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.
- Boye, Kasper & Peter Harder. 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in language* 31, 569-606.
- Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88:1, 1-44.
- Braunmüller, Kurt. 1998. *De nordiske språk*. [The Nordic languages] Oslo: Novus forlag.
- Brems, Lieselotte. 2007. The grammaticalization of small size nouns: Reconsidering frequency and analogy. *Journal of English Linguistics* 35, 293-325.
- Brems, Lieselotte, Kristin Davidse & Liesbeth De Smet. 2008. Type noun uses in the English NP: A case of right to left layering. *International Journal of Corpus Linguistics* 13, 139-168.
- Brinton, Laurel J. 1996. *Pragmatic markers in English; Grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter.
- Brinton, Laurel J. 2001. From matrix clause to pragmatic marker: The history of *look*-forms. *Journal of Historical Pragmatics* 2, 177-199.
- Brinton, Laurel J. 2008. *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge: Cambridge University Press.
- Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Bréal, Michel. 1964 [1900]. *Semantics: Studies in the Science of Meaning*. Cust, Henry (trans.). New York: Dover.
- Bussmann, Hadumod. 1996. *Routledge Dictionary of Language and Linguistics* (Trauth, Gregory & Kerstin Kazzazi (eds. and trans.) London/New York: Routledge.
- Bybee, Joan L. & William Pagliuca. 1985. Cross-linguistic comparison and the development of grammatical meaning. In Fisiak, J. (ed) *Historical semantics, historical word formation*, 59-83. The Hague: Mouton.
- Bybee, Joan L, William Pagliuca, Revere D. Perkins. 1991. Back to the future. In Traugott, Elizabeth Closs & Bernd Heine (eds) *Approaches to grammaticalization*, volume II, 17-58. Amsterdam: John Benjamins.
- Bybee, Joan L, William Pagliuca & Revere D. Perkins. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

- Campbell, Lyle & Richard D. Janda (eds). 2001. Grammaticalization: A critical assessment. Special issue. *Language Sciences* 23:2-3, 163-186.
- Campbell, Lyle & Richard Janda. 2001. Introduction: conceptions of grammaticalization and their problems. *Languages Sciences* 23, 93-112.
- Campbell, Lyle. 2001. What's wrong with grammaticalization? *Languages Sciences* 23, 113-161.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
- Crevels, Emily I. 2000. *Concession: a typological study*. Doctoral dissertation, University of Amsterdam.
- Davidse, Kristin, Lieven Vandelanotte & Hubert Cuyckens (eds) 2010. *Subjectification, intersubjectification and grammaticalization*. Berlin/New York: De Gruyter Mouton.
- Degand, Liesbeth & Anne-Marie Simon-Vandenberghe (eds). 2011. Grammaticalization, Pragmaticalization and/or (Inter)Subjectification: Methodological issues for the study of discourse markers. Thematic issue. *Linguistics* 49:2.
- de Haan, Ferdinand. 1999. Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics* 18, 83-101.
- de Haan, Ferdinand. 2006. Typological approaches to modality. In Frawley, William (ed) *The Expression of Modality*, 27-69. Berlin/New York: Mouton de Gruyter.
- de Haan, Ferdinand & Björn Hansen. 2009. Introduction. In Hansen, Björn & Ferdinand de Haan (eds) *Modals in the languages of Europe: a reference work*, 1-10. Berlin / New York: Mouton de Gruyter.
- Dehé, Nicole & Anne Wichmann. 2010. Sentence-initial I think (that) and I believe (that): Prosodic evidence for use as main clause, comment clause and discourse marker. *Studies in Language* 34:1, 36-74.
- Depraetere, Ilse & Susan Reed. 2006. Mood and Modality in English. In Aarts, Bas & April McMahon (eds) *The Handbook of English Linguistics*, 269-290. Malden: Blackwell Publishing Ltd.
- De Smet, Hendrik & Jean-Christophe Verstraete. 2006. Coming to terms with subjectivity. *Cognitive Linguistics* 17, 365-392.
- Detges, Ulrich & Richard Wältereit. 2002. Grammaticalization vs. Reanalysis: A Semantic-Pragmatic Account of Functional Change in Grammar. *Zeitschrift für Sprachwissenschaft* 21, 151-195.
- Detges, Ulrich & Richard Wältereit. 2009. Diachronic Pathways and Pragmatic Strategies: Different Types of Pragmatic Particles from a Diachronic Point of View. In Hansen, Maj-Britt Mosegaard & Jacqueline Visconti (eds) *Current Trends in Diachronic Semantics and Pragmatics*. Bingley: Emerald.
- Diderichsen, P. 1946. *Elementær Dansk Grammatik*. [Elementary Danish grammar] Copenhagen: Gyldendal.

- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49:2, 365-390.
- Doherty, Monika. 1987. *Epistemic Meaning*. Berlin etc: Springer Verlag.
- Dostie, Gaétane. 2009. Discourse markers and regional variation in French. A lexico-semantic approach. In Beeching, K., N. Armstrong & F. Gadet (eds) *Sociolinguistic variation in contemporary French*, 201-214. Amsterdam: John Benjamins.
- Eide, Kristin Melum. 2005. *Norwegian modals*. Berlin/New York: Mouton de Gruyter.
- Enger, Hans-Olav. in press. Morphological theory and grammaticalisation: the role of meaning and local generalisations. Special issue. *Language Sciences*.
- Erman, Britt & Ulla-Britt Kotsinas. 1993. Pragmaticalization: The case of ba' and you know. *Studier i modern språkvetenskap*, 76-93. Stockholm: Almqvist and Wiksell.
- Erteschik-Shir, Nomi. 2010. The phonology of adverb placement, OS and V-2: The case of Danish 'MON'. In Erteschik-Shir, Nomi & Lisa Rochman (eds) *The Sound Patterns of Syntax*, 31-35. Oxford: Oxford University Press.
- Evans, Nicholas. 2007. Insubordination and its Uses. In Nikolaeva (ed) *Finiteness*. Oxford: Oxford University Press.
- Evans, Nicholas D. & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception. *Language* 76, 546-592.
- Faarlund, Jan Terje. 2004. *The syntax of Old Norse*. Oxford: Oxford University Press.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. [Reference grammar of Norwegian.] Oslo: Universitetsforlaget.
- Falk, Hjalmar & Alf Torp. 1903. *Etymologisk ordbog over det norske og det danske sprog*. [Etymological dictionary of the Norwegian and Danish language.] Kristiana. <http://www.edd.uio.no/perl/search/search.cgi?appid=208&tabid=2320> (19-04-2012).
- Field, Andy. 2005. *Discovering statistics using SPSS*. London: Sage.
- Fillmore, Charles J. 1970. Types of lexical information. In Kiefer, F. (ed) *Studies in Syntax and Semantics*, 109-137. Dordrecht, Holland: D. Reidel.
- Fischer, Olga. 2007. The development of English parentheticals: A case of grammaticalization? In Smit, U. Dollinger, S. Hüttner, J., Kaltenböck, G., Lutzky, U. (eds) *Tracing English Through Time*, 99-114. Wien: Braumüller.
- Fischer, Olga, Anette Rosenbach & Dieter Stein (eds). 2000. *Pathways of change: grammaticalization in English*. Amsterdam: John Benjamins.
- Fischer, Olga & Annette Rosenbach. 2000. Introduction. In Fischer, Olga, Annette Rosenbach & Dieter Stein (eds) *Pathways of change: grammaticalization in English*, 1-37. Amsterdam/Philadelphia: John Benjamins.
- Fischer, Olga, Muriel Norde & Harry Perridon (eds). 2004. *Up and down the cline, the nature of grammaticalization*. Amsterdam: John Benjamins.
- Foley, William & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Fritzner, Johan. 1867. *Ordbog over Det gamle norske Sprog*. [Dictionary of the old Norwegian language] Kristiana: Den Norske Forlagsforening.



- Givón, T. 1979. *On Understanding Grammar*. New York: Academic Press.
- Gooskens, Charlotte, Karin Beijering & Wilbert Heeringa. 2008. Phonetic and lexical predictors of intelligibility. *International Journal of Humanities and Arts Computing*, 63-81.
- Goossens, Louis. 1982. On the development of the modals and of the epistemic function in English. Ahlqvist, Anders (ed) *Papers from the fifth International Conference on Historical Linguistics*, 74–84. Amsterdam/Philadelphia: John Benjamins.
- Gorell, J. Hendren. 1895. Indirect discourse in Anglo-Saxon. *Publications of the Modern Language Association* 10, 342-485.
- Greenberg, Joseph H. 1991. The last stages of grammatical elements: contractive and expansive desemanticization. In Traugott, Elizabeth Closs & Bernd Heine (eds) *Approaches to grammaticalization*, volume I, 301-314. Amsterdam: John Benjamins Publishing Company.
- Günthner, Susanne & Katrin Mutz. 2004. Grammaticalization vs. pragmaticalization? The development of pragmatic markers in German and Italian. In Bisang, Walter, Nikolaus P. Himmelmann & Björn Wiemer (eds) *What makes grammaticalization? A look from its fringes and its components*, 77-107. Berlin/New York: Mouton de Gruyter.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. Second edition. London: Arnold.
- Halliday, M. A. K & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Hansen, Maj-Britt Mosegaard. 1998. The semantic status of discourse markers. *Lingua* 104, 235-260.
- Hansen, Björn & Ferdinand de Haan (eds). 2009. *Modals in the languages of Europe: a reference work*. Berlin/New York: Mouton de Gruyter.
- Harris, A. & L. Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 1998. Does grammaticalization need reanalysis? *Studies in Language* 22, 49-85.
- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Fischer, Olga, Muriel Norde & Harry Perridon (eds) *Up and down the cline: the nature of grammaticalization*, 17–44. Amsterdam: John Benjamins.
- Heine, Bernd. 1993. *Auxiliaries: cognitive forces and grammaticalization*. Oxford: Oxford University Press.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Wisner, Ilse & Gabriele Diewald (eds) *New reflections on grammaticalization*, 83-101. Amsterdam: John Benjamins.
- Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Buske Verlag.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: a conceptual framework*. Chicago: The University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

- Hengeveld, Kees. 1988. Illocution, mood and modality in a functional grammar of Spanish. *Journal of Semantics* 6, 227–269.
- Himmelman, Nikolaus P. 2004. Lexicalization and grammaticalization: Opposite or orthogonal? In Bisang, Walter, Nikolaus P. Himmelman & Björn Wiemer (eds) *What Makes Grammaticalization? A Look from its Fringes and its Components*, 21-42. Berlin/New York: Mouton de Gruyter.
- Holmberg, Anders & Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. New York: Oxford University Press.
- Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Second edition. Cambridge: Cambridge University Press.
- Hopper, Paul J. 1991. On some principles of grammaticization. In Traugott, Elizabeth Closs & Bernd Heine (eds) *Approaches to grammaticalization*, volume I, 17-35. Amsterdam: John Benjamins.
- Hoye, Leo. 1997. *Adverbs and Modality in English*. London/New York: Longman.
- Jakobson, Roman & Morris Halle. 1956. *Fundamentals of Language*. The Hague: Mouton.
- Joseph, Brian D. 2005. How accommodating of change is grammaticalization? The case of “lateral shifts”. *Logos and Language* 6, 1-7.
- Kaufmann, Stefan, Cleo Condoravdi & Valentina Harizanov. 2006. Formal approaches to modality. In Frawley, William (ed) *The expression of modality*, 71-106. Berlin/New York: Mouton de Gruyter.
- Kearns, Kate. 2007. Epistemic verbs and zero complementizer. *English Language and Linguistics* 11, 475-505.
- Keski-Raasakka, Kyllikki. 2002. Männe Lisa är hemma? Männe inte? Om männe i svenska skolgrammatikor och några korpusar. *Svenskans beskrivning* 25. Åbo: Åbo universitet.
- Kiparsky, Paul & Carol Kiparsky. 1970. Fact. In M. Bierwisch & K. Heidolph (eds.) *Progress in Linguistics*. The Hague: Mouton.
- Kuryłowicz, Jerzy. 1975 [1965]. The evolution of grammatical categories. *Esquisses linguistiques* 2, 38-54.
- Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. *Diogenes*, 55-71.
- Kuteva, Tania. 2001. *Auxiliation: an enquiry into the nature of grammaticalization*. New York/Oxford: Oxford University Press.
- Langacker, Ronald W. 1985. Observations and speculations on subjectivity. In Haiman, John (ed) *Iconicity in Syntax*, 109-150. Amsterdam: John Benjamins.
- Langacker, Ronald W. 1990. Subjectification. *Cognitive Linguistics* 1, 5-38.
- Lass, Roger. 1990. How to do things with junk: exaptation in language evolution. *Journal of Linguistics* 26, 79–102.
- Lehmann, Christian. 1982. *Thoughts on grammaticalization: a Programmatic Sketch*. Cologne: University of Cologne, Institut für Sprachwissenschaft.

- Lehmann, Christian. 1995. *Thoughts on grammaticalization*. München: LINCOM Europa.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In Wischer, Ilse & Gabriele Diewald (eds) *New reflections on grammaticalization*, 1-18. Amsterdam: John Benjamins.
- Lenker, Ursula. 2000. *Soþlice* and *witodlice* discourse markers in Old English. In Fischer, Olga, Anette Rosenbach & Dieter Stein (eds) *Pathways of change: grammaticalization in English*, 229-249. Amsterdam/Philadelphia: John Benjamins.
- Lightfoot, David W. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lindström, Therese. 2004. The history of the Concept of Grammaticalisation. PhD Thesis, University of Sheffield. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-55932> (16-04-2012).
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Lyons, John. 1982. Deixis and subjectivity: *Loquor, ergo sum?* In Jarvella, Robert J. & Wolfgang Klein (eds) *Speech, Place, and Action: Studies in Deixis and Related Topics*, 101-124. New York: Wiley.
- Matisoff, James A. 1991. Areal and universal dimensions of grammatization in Lahu. In Traugott, Elizabeth Closs & Bernd Heine (eds) *Approaches to grammaticalization*, volume II, 383-453. Amsterdam: John Benjamins.
- McCawley, James D. 1968. Lexical insertion in a transformational grammar without deep structure. In Darden, Bill J., Charles-James N. Bailey & Alice Davison (eds) *Papers from the Fourth Regional Meeting, Chicago Linguistic Society*, 71-80. Chicago: Chicago Linguistic Society.
- Meillet, Antoine. 1912 [1958]. L'évolution des formes grammaticales. *Scientia* 12.
- Meillet, Antoine. 1958. *Linguistique historique et linguistique générale*, 1:130-48. Paris: Edouard Champion.
- Mortelmans, Tanja, Kasper Boye & Johan van der Auwera. 2009. Modals in the Germanic languages. In Hansen, Björn & Ferdinand de Haan (eds) *Modals in the languages of Europe: a reference work*. Berlin/New York: Mouton de Gruyter.
- Narrog, Heiko & Bernd Heine (eds). 2011. *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.
- Newmeyer, Frederick J. 1998. *Language Form and Language Function*. Cambridge, MA: MIT Press, Bradford Books.
- Newmeyer, Frederick J. 2001. Deconstructing grammaticalization. *Language Sciences* 23, 187-220.
- Norde, Muriel. 2006a. Demarcating degrammaticalization: the Swedish s-genitive revisited. *Nordic Journal of Linguistics* 29:2, 201-238.
- Norde, Muriel. 2006b. Van suffix tot telwoord tot bijwoord: degrammaticalisering en (re)grammaticalisering van *tig*. *TABU* 35:1/2, 33-60.
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

- Norde, Muriel. to appear. Lehmann's parameters revisited. In Breban, Tine, Lot Brems, Kristin Davidse & Tanja Mortelmans (eds) *Grammaticalization and Language Change: origins, criteria and outcomes*. Amsterdam/Philadelphia: John Benjamins.
- Norde, Muriel. 2012. Corpus-based approaches to constructional change in progress: the case of semi-autonomous subordination in Swedish. NIAS-Constructional change in the languages of Europe, Wassenaar, January 12-14 2012.
- Norde, Muriel & Karin Beijering. 2012. Facing interfaces – a clustering approach. Paper presented at Refining Grammaticalization workshop, Berlin, February 24-25 2012.
- Norde, Muriel, Gudrun Rawoens & Karin Beijering. 2012. En diakron studie av det epistemiska satsadverbet *kanske*. Paper presented at *Svenska språkets historia 12*, April 13-14 2012, Stockholm. <http://www.nordiska.su.se/pub/jsp/polopoly.jsp?d=16696> (20-04-2012).
- Norde, Muriel, Karin Beijering & Gudrun Rawoens. 2011. From matrix to sentence adverb or vice versa? The history of Swedish *kanske* 'maybe.' International workshop on Gradualness in change and its relation to synchronic variation and use, Pavia, May 30-31 2011.
- Norde, Muriel, Karin Beijering & Alexandra Lenz (eds). in press. Current trends in grammaticalization research. Special issue. *Language Sciences*.
- Nuyts, Jan. 2000. Tensions between discourse structure and conceptual semantics: The syntax of epistemic modal expressions. *Studies in Language* 24, 103-135.
- Nuyts, Jan. 2001. Epistemic Modality, Language, and Conceptualization. A Cognitive Pragmatic Perspective. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Nuyts, Jan. 2001a. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33, 383-400.
- Nuyts, Jan. 2005. The modal confusion: On terminology and the concepts behind it. In Klinge, A & H. H. Müller (eds) *Modality: Studies in form and function*, 5-38. London: Equinox.
- Nuyts, Jan. 2006. Modality: Overview and linguistic issues. In Frawley, William (ed) *The expression of modality*, 1-26. Berlin/New York: Mouton de Gruyter.
- Nuyts, Jan. forthc. De-auxiliarization without de-modalization in the Dutch core modals: A case of collective degrammaticalization? *Language Sciences*. In print. Doi 10.1016/j.langsci.2012.04.003
- Nuyts, Jan, Pieter Byloo and Janneke Diepeveen. 2010. On deontic modality, directivity, and mood: The case of Dutch *mogen* and *moeten*. *Journal of Pragmatics* 42, 16-34.
- Nørgård-Sørensen, Jens, Lars Heltoft & Lene Schøsler. 2011. *Connecting Grammaticalisation*. Amsterdam: John Benjamins.
- Ocampo, Francisco. 2006. Movement towards discourse is not grammaticalization: the evolution of *claro* from adjective to discourse particle in spoken Spanish. In Sagarra, Nuria & Almeida Jacqueline Toribio (eds) *Selected proceedings of the 9<sup>th</sup> Hispanic Linguistics Symposium*, 308-319. Somerville, MA: Cascadilla Proceedings Project.
- Onodera, Noriko. 1995. Diachronic analysis of Japanese discourse markers. In Andrea Jucker (ed) *Historical Pragmatics*, 393-437. Amsterdam: John Benjamins.
- Palmer, Frank R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.

- Palmer, Frank R. 2001. *Mood and modality*. Second edition. Cambridge: Cambridge University Press.
- Perkins, Michael R. 1983. *Modal Expressions in English*. London: Frances Pinter (Publishers).
- Petersson, David & Gunlög Josefsson. 2010. ELLERHUR and other Yes/No-question operator candidates in Swedish. *Working Papers in Scandinavian Syntax* 86, 179-198.
- Philippa, Marlies, Frans Debrabandere, Arend Quak, Tanneke Schoonheim & Nicoline van der Sijs (eds). 2011. *Etymologisch woordenboek van het Nederlands*. [Etymological dictionary of the Dutch language]. Amsterdam: Amsterdam University Press. <http://www.etymologie.nl/> (19-04-2012)
- Pinto de Lima, José. 2002. Grammaticalization, subjectification and the origin of phatic markers. In Wischer, Ilse & Gabriele Diewald (eds) *New reflections on grammaticalization*, 363-378. Amsterdam/Philadelphia: John Benjamins.
- Pintzuk, Susan. 1991. *Phrase Structures in Competition: Variation and Change in Old English Word Order*. Dissertation, University of Pennsylvania, Pittsburg.
- Platzack, C. 1998. *Svenskans inre grammatik – det minimalistiska programmet. En introduktion till modern generativ grammatik*. [The inner grammar of Swedish-the Minimalist Program. An introduction to modern generative grammar.] Lund: Studentlitteratur.
- Platzack, C. 1999. Review of Norsk referansegrammatikk. *Maal og Minne* 1, 91-108.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Ramat, Paolo. 2001. Degrammaticalization or transcategorization? In Schaner-Wolles, Chris, John Rennison & Friedrich Neubarth (eds) *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of his 60th Birthday*, 393-401. Torino: Rosenbach and Sellier.
- Ramat, Paolo & Davide Ricca. 1994. Prototypical adverbs: On the scalarity/radiality of the notion of ADVERB. *Rivista di Linguistica* 6, 289-326.
- Ramat, Paolo & Davide Ricca. 1998. Sentence adverbs in the languages of Europe. In van der Auwera, Johan & Dónall P. Ó Baoill (eds) *Adverbial constructions in the languages of Europe*, 187-273. Berlin/New York: Mouton de Gruyter.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.
- Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.
- Rosenkvist, Henrik. 2004. *The Emergence of Conditional Subordinators in Swedish – a Study in Grammaticalization*. Lund: Studentlitteratur.
- Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- Sidnell, Jack. 2009. Deixis. In Verschueren, Jef & Jan-Ola Östman (eds) *Key Notions for Pragmatics*. Amsterdam/Philadelphia: John Benjamins.
- Stubbs, M. 1983. *Discourse analysis: The sociolinguistic analysis of natural language*. Chicago: University of Chicago Press.

- Sundman, Marketta. 1983. *Svenska modalverb - ett kontinuum från hjälpverb till huvudverb?* [Modal verbs in Swedish – a continuum from auxiliary to main verb?] In Andersson Erik, Mirja Saari & Peter Slotte (eds) *Struktur och variation* (= meddelanden från Stiftelsen för Åbo akademi forskningsinstitut 85). Åbo: Åbo Akademi.
- Sweetser, Eve. 1982. Root and epistemic modals: causality in two worlds. In Macaulay, Monica, Orin D. Gensler et al. (eds) *Proceedings of the eighth Annual Meeting of the Berkeley Linguistics Society*, 484–507. Berkeley, CA: Berkeley Linguistics Society.
- Tabor, Whitney & Elizabeth Closs Traugott. 1998. Structural scope expansion and grammaticalization. In Ramat, Anna Giacalone & Paul J. Hopper (eds) *The limits of grammaticalization*, 229-272. Amsterdam: John Benjamins.
- Teleman, Ulf. 1974. *Manual för grammatisk beskrivning av talad och skriven svenska*. [Manual for the description of spoken and written Swedish.] Lund: Studentlitteratur.
- Teleman, Ulf, Staffan Hellberg, Erik Andersson & Lisa Christensen (eds). 1999. *Svenska Akademiens Grammatik*. [Grammar of the Swedish Academy] Stockholm: Svenska Akademien/Norstedts Ordbok.
- Thompson, Sandra A. & Anthony Mulac. 1991. A quantitative perspective on the grammaticalization of epistemic parentheticals in English. In Traugott, Elizabeth Closs & Bernd Heine (eds) *Approaches to Grammaticalization*, 313-329. Amsterdam/Philadelphia: John Benjamins.
- Torp, Arne. 1998. *Nordiske språk i nordisk og germansk perspektiv* [Nordic languages in Nordic and Germanic perspective]. Oslo: Novus forlag.
- Traugott, Elizabeth Closs. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Lehmann, Winfred P. & Yakov Malkiel (eds) *Perspectives on Historical Linguistics*, 245-271. Amsterdam/Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* 65:1, 31-55.
- Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalisation. In Stein, Dieter & Susan Wright (eds) *Subjectivity and subjectivisation*, 31-54. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs. 2002. From etymology to historical pragmatics. In D. Minkova & R. Stockwell (eds) *Studies in the history of the English language*, 19–49. Berlin: Mouton de Gruyter.
- Traugott, Elizabeth Closs. 2003. From subjectification to intersubjectification. In Hickey, Raymond (ed) *Motives for Language Change*, 124–139. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs. 2001. *Legitimate counterexamples to unidirectionality*. Paper presented at Freiburg University, October 17th 2001. <http://www.stanford.edu/~traugott/papers/Freiburg.Unidirect.pdf> (17-04-2012).

- Traugott, Elizabeth Closs. 2010. (Inter)subjectivity and (inter)subjectification. In Davidse, Kristin, Lieven Vandelanotte & Hubert Cuyckens (eds) *Subjectification, intersubjectification and grammaticalization*, 29-71. Berlin/New York: De Gruyter Mouton.
- Traugott, Elizabeth Closs. 2010a. Grammaticalization. In Luraghi, Silvia & Vit Bubenik (eds) *The Continuum Companion to Historical Linguistics*, 271-285. London: Continuum Press.
- Traugott, Elizabeth Closs & Bernd Heine. 1991. *Approaches to grammaticalization*, volume I and II. Amsterdam: John Benjamins.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2010. *Gradience, gradualness and grammaticalization*. Amsterdam: John Benjamins.
- Trousdale, Graeme. forthc. Grammaticalization, constructions and the grammaticalization of constructions. In Breban, Tine, Lot Brems, Kristin Davidse & Tanja Mortelmans (eds) *Grammaticalization and Language Change: Origins, Criteria and Outcomes*. Amsterdam/Philadelphia: John Benjamins.
- Urmson, James O. 1952. Parenthetical verbs. *Mind* 61, 480-496.
- Van Bogaert, Julie. 2010. A constructional taxonomy of I think and related expressions: accounting for the variability of complement-taking mental predicates. *English Language and Linguistics* 14:3, 399-427.
- van der Auwera, Johan. 2001. On the typology of negative modals. In Jack Hoeksema, Hotze Rullmann, Victor Sanchez-Valencia & Ton van der Wouden (eds) *Perspectives on negation and polarity items*, 23-48. Amsterdam: Benjamins.
- van der Auwera, Johan. 2002. More thoughts on degrammaticalization. In Wischer, Ilse & Gabriele Diewald (eds) *New reflections on grammaticalization*, 19-29. Amsterdam: John Benjamins.
- van der Auwera, Johan & Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2, 79-124.
- van der Auwera, Johan, Petar Kehayov & Alice Vittrant. 2009. Acquisitive modals. In Hogeweg, L., H. De Hoop & A. Malchukov. *Cross-linguistic semantics of tense, aspect and modality*, 271-302. Amsterdam: Benjamins.
- Vande Velde, Freek & An Van linden. 2010. (Semi-)autonomous subordination patterns in Dutch: A unified account. Coglingdagen. Antwerpen, 17-18 december 2010.
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam/Philadelphia: John Benjamins.
- Vikner, S. 1995. *Verb Movement and Expletive Subjects in the Germanic languages*. New York: Oxford University Press.
- Visconti, Jacqueline. forthc. Facets of subjectification. *Language Sciences*.
- Von Mengden, Ferdinand. 2008. The modules of grammatical change. Paper presented at *New reflections on grammaticalization 4*, Leuven, July 16<sup>th</sup>-19<sup>th</sup>, 2008.

- von Mengden, Ferdinand & Horst Simon. 2012. Description of the Workshop.  
[http://www.geisteswissenschaften.fuberlin.de/en/v/Refining\\_Grammaticalization/index.html](http://www.geisteswissenschaften.fuberlin.de/en/v/Refining_Grammaticalization/index.html)  
 (16-04-2012)
- von Wright, Georg Henrik. 1951. *An essay in modal logic*. Amsterdam: North-Holland.
- Wessén, Elias. 1965. *Svensk språkhistoria. I Ljudlära och ordböjningslära*. [History of the Swedish language: I phonetics and morphology] Stockholm: Almqvist & Wiksell.
- Wessén, Elias. 1967. Ett fornsvenskt vardagsord Fsv. maxan – da. måske – sv. kanske. [An Old Swedish common word Old Swedish maxan - Danish måske – Swedish kanske] *Nysvenska Studier* 47. Lund: Carl Bloms Boktryckeri A.-B.
- Wischer, Ilse. 2000. Grammaticalization versus lexicalization – “methinks” there is some confusion. In Fischer, Olga, Anette Rosenbach & Dieter Stein (eds) *Pathways of change: grammaticalization in English*, 355-370. Amsterdam: John Benjamins Publishing Company.
- Wischer, Ilse & Gabriele Diewald (eds). 2002. *New reflections on grammaticalization*. Amsterdam: John Benjamins Publishing Company.
- Zola Christensen, R. & L. Christensen. 2005. *Dansk grammatik*. Odense: Syddansk Universitetsforlag.



## Summary

This dissertation investigates the rise of epistemic expressions in relation to different types of language change, viz. lexicalization, grammaticalization and pragmaticalization. Four case studies from the domain of epistemic modality in the Mainland Scandinavian languages, i.e. Danish, Norwegian and Swedish, have been discussed. These case studies are concerned with the modal auxiliaries *MUST/MAY* (Chapter 3), the interrogative/epistemic adverb *MON* ‘I wonder’, the epistemic adverb *MAYBE* (Chapter 4), and the discourse marker *I THINK* (Chapter 5). These epistemic phenomena pose problems for traditional analyses within the framework of grammaticalization studies, because they are all at the interface of (most current definitions of) grammaticalization, lexicalization and pragmaticalization.

Chapter 1 introduced and motivated the topic of this dissertation. The concept of epistemic modality and related notions has been discussed and defined within the broader domain of modality. The Scandinavian languages, their interrelations and history have been described and the framework of grammaticalization studies has been discussed from a functional and formal perspective.

In Chapter 2 various definitions and descriptions of lexicalization, grammaticalization, pragmaticalization and (inter)subjectification have been reviewed. The notions of lexicalization, grammaticalization and pragmaticalization have been redefined in a unified model of language change. The converging and diverging properties of these different types of language determine the unique properties of lexicalization, grammaticalization and pragmaticalization, as well as their shared properties, i.e. interfaces.

Chapter 3 presented the first case study on the modal auxiliaries *MUST/MAY*. This chapter focused on the differences and similarities between Danish, Norwegian and Swedish with respect to the semantic distributions and formal properties of the modals *MUST/MAY* and the interaction between grammaticalization and (inter)subjectification in the development of modal and postmodal meanings. Significant differences, both formally and semantically, have been found with respect to the development of Mainland Scandinavian *MUST/MAY*. Their development is in line with observed developmental tendencies for their Germanic cognates. That is, these modals have the ability to express various modal and postmodal dimensions and are moving towards auxiliary status.

The second case study on the modal adverbs *MON* and *MAYBE* is described in Chapter 4. In this chapter the status of intra-categorical changes, the lexicalization-grammaticalization-pragmaticalization interface and the peculiar syntactic status of *MON* and *MAYBE* have been discussed. Both *MON* and *MAYBE* display remarkable syntactic behaviour. They may violate the V2- principle, which can be explained by taking into account their historical development and verbal origin. The results of the case study on Mainland Scandinavian *MON* highlight considerable differences, both formally and semantically, with respect to its status and development. The interrogative adverb *MON* expresses various dimensions of doubt from the point of view of the speaker and may occur in different syntactic constructions. The general picture is that Danish, Norwegian and Swedish developed their own idiosyncrasies and syntactic

structures for MON. The results of the case study on Mainland Scandinavian MAYBE highlight substantial differences with respect to its syntactic distribution. All in all, MAYBE functions more and more as a pure sentence adverb in V2 clauses. MAYBE is losing its subordinating properties and non-V2 clauses are gradually being replaced by V2 clauses.

Chapter 5 presented the third case study on Mainland Scandinavian I THINK. This chapter described the differences and similarities for I THINK with respect to its etymology, development and syntactic and semantic properties. The focus of this chapter is the development of the discourse marker I THINK in relation to the lexicalization-grammaticalization-pragmaticalization interface and (inter)subjectification. The general picture that arises from the random samples is that there is a lot of syntactic and semantic variation for I THINK, which suggests that this mental state predicate is in full development. It is concluded that the complement taking predicate I THINK is an instance of subjectification and the discourse marker I THINK is an instance of pragmaticalization.

Chapter 6 introduced a novel model of analysis and a unified account of language change centered on the notion of a composite change. A composite change is composed of formal reanalysis and semantic reinterpretation, accompanying primitive changes at different linguistic levels (phonology, morphology, syntac, semantics, discourse/pragmatics), and the concomitants of these changes. These (micro-)changes cluster with respect to their converging properties (=interfaces), but stand out in case they are diverging properties (=unique properties). This is the essence of the clustering approach to language change. It has been shown that primitive changes tend to form clusters in terms of reduction or expansion (e.g. semantic bleaching versus enrichment, morphological fusion versus separation) or no change at all. These clusters may correspond to changes traditionally labeled 'grammaticalization', 'lexicalization' or 'pragmaticalization', but changes may also cluster in alternative ways. This is what happened in the case studies in this dissertation.

The case studies in this dissertation defy strict categorization. Therefore, it is proposed to give up the idea that linguistic changes can be assigned to predefined categories. Instead, it has been argued that it is more sensible to reduce lexicalization, grammaticalization and pragmaticalization to their basic mechanisms of change, primitive changes and the concomitants of these (micro-)changes. The advantage of clustering, as opposed to categorization, is that one can identify prototypical and marginal instances of a certain type of language change. Moreover, the clustering approach indicates the interface areas between lexicalization, grammaticalization and pragmaticalization. Because of this, it is capable of accommodating hybrid or problematic items into the unified model of language change.

## Samenvatting

In dit proefschrift wordt een vergelijkende corpusstudie naar het ontstaan van epistemische uitdrukkingen beschreven en geanalyseerd in relatie tot verschillende typen taalverandering, te weten lexicalisatie, grammaticalisatie en pragmaticalisatie. Vier casestudies uit het domein van epistemische modaliteit in de continentaal-Scandinavische talen, i.e. Deens, Noors en Zweeds, worden besproken. De case studies hebben betrekking op het modale (hulp)werkwoord MOETEN/MOGEN (Hoofdstuk 3), het epistemisch/vraagpartikel MON ‘ik vraag me af’, het epistemische bijwoord MISSCHIEN (Hoofdstuk 4), en de discourse marker IK DENK (Hoofdstuk 5). Deze epistemische uitdrukkingen vormen een probleem voor traditionele analyses binnen grammaticalisatie studies omdat ze zich in het grensgebied tussen deze verschillende typen taalverandering bevinden.

In hoofdstuk 1 wordt het onderwerp van deze dissertatie geïntroduceerd en toegelicht. Epistemische modaliteit en verwante begrippen worden besproken en gedefinieerd binnen het overkoepelende domein van modaliteit. De geschiedenis en onderlinge relaties tussen de continentaal-Scandinavische talen worden beschreven. Het theoretisch kader waarbinnen het onderzoek is uitgevoerd, grammaticalisatie studies, wordt zowel vanuit functioneel als generatief perspectief belicht.

In hoofdstuk 2 worden verschillende definities en beschrijvingen van lexicalisatie, grammaticalisatie, pragmaticalisatie en (inter)subjectificatie besproken en vergeleken. De kernbegrippen lexicalisatie, grammaticalisatie en pragmaticalisatie worden geherdefinieerd in een verenigd model van taalverandering. De convergerende en divergerende eigenschappen van de verschillende typen taalverandering bepalen de unieke kenmerken van lexicalisatie, grammaticalisatie en pragmaticalisatie, maar ook de grensgebieden tussen deze typen taalverandering.

Hoofdstuk 3 beschrijft de eerste casestudy naar het modale (hulp)werkwoord MOETEN/MOGEN. De verschillen en overeenkomsten wat betreft de semantische distributie en formele eigenschappen van MOETEN/MOGEN en de interactie tussen grammaticalisatie en (inter)subjectificatie in het ontstaan van modale en postmodale betekenissen staan centraal in dit hoofdstuk. Er zijn significante verschillen, zowel formeel als semantisch, in de ontwikkeling van deze modalen in het Deens, Noors en Zweeds. De ontwikkeling van MOETEN/MOGEN volgt dezelfde welbekende tendensen in het ontstaan van modalen als voor cognaat modalen in andere Germaanse talen. Oftewel, deze modalen kunnen verschillende modale en postmodale betekenissen uitdrukken en transformeren geleidelijkaan van lexicaal werkwoord naar hulpwerkwoord.

De tweede casestudy wordt besproken in hoofdstuk 4 en heeft betrekking op de modale adverbia MON ‘ik vraag me af’ en MISSCHIEN. Dit hoofdstuk concentreert zich met name op de status van intra-categoriale veranderingen, de overlap tussen lexicalisatie, grammaticalisatie en pragmaticalisatie, en de bijzondere syntactische eigenschappen van MON en MISSCHIEN. Zowel MON als MISSCHIEN kunnen het V2-principe schenden in declaratieve hoofdzinnen. Dit wordt verklaard door de verbale oorsprong en historische ontwikkeling van MON en MISSCHIEN. De

resultaten van de casestudy naar MON laat aanzienlijke verschillen zien, zowel formeel als semantisch, wat betreft de status en ontwikkeling van MON. In vraagzinnen voegt het modale vraagpartikel MON een extra dimensie van persoonlijke twijfel toe. Er zijn verschillende syntactische constructies waar MON deel van uit kan maken. Het algemene beeld is dat MON in het Deens, Noors en Zweeds taalspecifieke eigenschappen en syntactische structuren heeft. De resultaten van de casestudy naar MISSCHIEN laat substantiële verschillen in de syntactische distributies zien. Al met al functioneert MISSCHIEN voornamelijk als zinsadverbium in V2 zinnen. MISSCHIEN verliest de onderschikkende eigenschappen en niet-V2 structuren worden geleidelijk aan vervangen door V2 zinnen.

Hoofdstuk 5 behandelt de derde casestudy naar de discourse marker IK DENK. De verschillen en overeenkomsten wat betreft de etymologie, ontwikkeling en semantische en formele eigenschappen van IK DENK worden beschreven voor het Deens, Noors en Zweeds. De focus van dit hoofdstuk is de totstandkoming van de discourse marker IK DENK in relatie tot (inter)subjectificatie en de grensgebieden tussen lexicalisatie, grammaticalisatie en pragmaticalisatie. De resultaten van de analyse van de corpusdata laten zien dat er veel syntactische en semantische variatie is. Dit suggereert dat IK DENK in volle ontwikkeling is. Het mentale predicaat IK DENK (DAT), met complement structuur, is een geval van subjectificatie, de discourse marker IK DENK, met parenthetische structuur, is een voorbeeld van pragmaticalisatie.

In hoofdstuk 6 wordt een nieuw analysemodel gepresenteerd dat is gebaseerd op het concept van de samengestelde verandering. In dit proefschrift worden lexicalisatie, grammaticalisatie en pragmaticalisatie gedefinieerd als samengestelde veranderingen die bestaan uit i) formele heranalyse en semantische herinterpretatie, ii) primitieve veranderingen op verschillende taalkundige niveaus (morfologie, fonologie, syntaxis, semantiek en pragmatiek), en iii) de bijeffecten van deze (micro-)veranderingen zoals bijvoorbeeld frequentie en productiviteit. Door de verschillende typen taalverandering op dezelfde manier te definiëren kunnen ze vergeleken worden in een verenigd model van taalverandering waarin de gemeenschappelijke en onderscheidende kenmerken zijn ondergebracht.

De analyse is gebaseerd op clusters van eigenschappen in plaats van strikte categorisatie. De clusters kunnen samenvallen met alle kenmerken die traditioneel geassocieerd worden met bijvoorbeeld grammaticalisatie, maar ze kunnen ook alternatieve patronen vormen en eigenschappen van meerdere typen verandering hebben. Dit is wat er gebeurt in de casestudies die besproken worden in dit proefschrift. De casestudies in dit proefschrift kunnen niet zonder meer geassocieerd worden als zijnde (prototypische) gevallen van lexicalisatie, grammaticalisatie of pragmaticalisatie. Daarom wordt het idee dat een verandering toegeschreven kan worden aan één bepaald type taalverandering losgelaten. In plaats daarvan wordt gekeken naar clusters van (micro-)veranderingen die een verandering karakteriseren zonder het meteen als lexicalisatie, grammaticalisatie of pragmaticalisatie te bestempelen. Het voordeel van de clustermethode ten opzichte van strikte categorisatie is dat prototypische en marginale gevallen van een bepaald type taalverandering kunnen worden onderscheiden, maar ook de grensgebieden tussen lexicalisatie, grammaticalisatie en pragmaticalisatie.

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorical Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.
12. Ton van der Wouden (1994). *Negative Contexts*.
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorical Grammar*.
14. Petra Hendriks (1995). *Comparatives and Categorical Grammar*.
15. Maarten de Wind (1995). *Inversion in French*.
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du' NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*.
29. H. Wee (1999). *Definite Focus*.
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse*.
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing*.
32. Klarien van der Linde (2001). *Sonority substitutions*.
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement*.
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*.
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study*.
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection*.

38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren*.
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits*.
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension*.
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition*.
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study*.
43. Hein van Schie (2003). *Visual Semantics*.
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian*.
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures*.
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*.
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology*.
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia*.
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision*.
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation*.
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin*.
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability*.
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis*.
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse*.
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency*.
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs*.
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality*.
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing*.
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music*.
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology*.
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing*.
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb*.
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism*.
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing*.
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting*.
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia*.
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch*.
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering*.

70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering*.
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian*.
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System*.
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case*.
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia*.
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition*.
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities*.
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment*.
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines*.
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval*.
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg*.
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships*.
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text*.
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*.
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning*.
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development*.
86. Teodora Mehotchewa (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student*.
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students*.
88. Jelena Prokić (2010). *Families and Resemblances*.
89. Radek Šimik (2011). *Modal existential wh-constructions*.
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease*.
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching*.
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters*.
93. Marlies Kluck (2011). *Sentence amalgamation*.
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*.
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation*.
96. Barbara Plank (2011). *Domain Adaptation for Parsing*.

97. Cagri Coltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*
99. Herman Heringa (2012). *Appositional constructions.*
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia.*
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation.*
104. Gisi Cannizzaro (2012). *Early word order and animacy.*
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model.*
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*

GRODIL

Center for Language and Cognition Groningen (CLCG)

P.O. Box 716

9700 AS Groningen

The Netherlands



